Order Number 8923854

Evaluation of multilevel sampling techniques for forest inventory in northern Michigan

Humphreys, Rubens Dias, Ph.D.

Michigan State University, 1989

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

EVALUATION OF MULTILEVEL SAMPLING TECHNIQUES FOR FOREST INVENTORY IN NORTHERN MICHIGAN

By

Rubens Dias Humphreys

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Forestry

1989

ABSTRACT

EVALUATION OF MULTILEVEL SAMPLING TECHNIQUES FOR FOREST INVENTORY IN NORTHERN MICHIGAN

By

Rubens Dias Humphreys

There are several sampling methods that can be applied to forest inventory. The selection of the most appropriate one depends on several aspects : number of forest type(s) in the area; size of the area; objectives of the inventory; funds available; availability of remotely sensed imagery, etc. In large areas, multilevel sampling techniques such as multistage or multiphase can be used. Incorporation of remotely sensed imagery is essential for the application of such methods.

Two multilevel sampling techniques were evaluated: a multistage method which has been used quite frequently in inventories in the USA and a multiphase technique which has been recently developed. The former was composed of three stages. The first stage used LANDSAT TM imagery enlarged to 1:107,000; the second stage utilized 1:24,000 CIR photographs and the third stage incorporated ground measurements. The later method was a stratified two phase technique where stratification of the area was done by

computer classification of the digital representation of the imagery. The first phase of this approach was a simple random sample within each stratum; the second phase was a sub-sample of the first, where field measurements were done. The study was conducted on an area of 47,850 acres in Wexford County, Michigan. Field measurements were done using point sampling and the volumes of trees were estimated using specific volume equations. Although using a reduced sample size, the sampling error for the multiphase technique was lower than that for the multistage technique. The size variables used to calculate the selection probabilities for the second stage sampling units (estimated area of the forest type contained on the secondary unit) and for the third stage sampling units (percent crown closure), were not appropriate. The size of the primary sampling units chosen was not convenient because: a) only two PSU's were contained within the red oak stratum; b) it was too large to be used for subdivision of the conifer stratum and c) produced variations on the selection probabilities calculated for the second stage, which resulted in negative correlations between its size variable and predicted volume.

This work is specially dedicated to my
parents, Brigida D. Humphreys and
Dr. Jorge Humphreys (in memoriam)

ACKNOWLEDGMENTS

I would like to express my appreciation to the Brazilian Government through the Ministerio da Educacao e Cultura for the financial support of my studies at Michigan State University. Also to the Instituto de Pesquisas Tecnologicas do Estado de Sao Paulo SA - IPT for allowing me to leave my duties in order to pursue my degree.

My appreciation and respect goes to my major professor Dr. Carl W. Ramm for his support, patience and guidance during all the steps of my graduate program and to my graduate committee Dr. Robert Marty, Dr. David P. Lusch and Dr. Karen Potter Witter for their guidance and constructive critiques of my work.

My most sincere appreciation to Geraldo Jose Zenid, Marcelo de Andrade and Joyce de Andrade for their invaluable support in taking care of our things; to Mr. Sergey G. Guins and his wife Mrs. Katherine Guins for their love and help in hours we needed most. To other friends from the Wood Division that directly or indirectly gave their contribution and support, even if small, but that sure ends up being significant.

To the Department of Natural Resources of Michigan and to Dr. Dennis W. Hudson from the Center for Remote Sensing, for allowing me to use the imagery of the study area.

Finally, but not the least important, my deepest gratitude to my wife Berenice for her love, encouragement and support during this laborious journey and to my daughter Sheila, although still a child but with a great capability of understanding.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The term inventory is defined by the Webster's Third International Dictionary of the English Language as: "an itemized list of current assets". The term was initially applied to commercial enterprises where it was necessary to control the stock of raw materials and the items produced.

The term forest is defined by FAO as : "all lands bearing a vegetation association dominated by trees of any size, exploited or not, capable of producing wood or other products, of exerting an influence on the climate or on the water regime, or providing shelter for lifestock and wild-life (Loetsch *et al*. 1964)".

By combining both words we form the term forest inventory. The "current assets" generally refers to the tree species that occur in a specific area of interest. The "itemized list" refers to the tabulated information about the trees. This information can be presented in several ways such as by tree species, diameter classes, quality. The information normally refers to estimated volume and to the number of trees occurring in the area, with an associated precision. The information may be expressed on a per unit are basis(hectare or acre), and the precision is a measure of its reliability.

The information to be obtained by a forest inventory depends on the objectives of the inventory (Loetsch *et al.* 1964; Nyyssonen, 1976). If the objective of the inventory is, for instance, future land use planning, recreational studies or watershed management, there is no need to gather information related to tree volume (Nyyssonen, 1976).

If, for instance, the aim is to identify projects for future implementation on the area, a reconnaissance inventory is appropriate. This type of inventory is done with the purpose of estimating total volume of standing trees, specifying volume by species, diameter classes and tree quality. Normally no emphasis is given to the uses of the available stock (Nyyssonen, 1976, 1978). The information gathered on reconnaissance inventories is of little use in making investment decisions.

Information for this objective should be obtained from a pre-investment inventory. The purpose of this type of inventory is to provide an answer to the following question (Nyyssonen, 1978): "How much industrial wood, specified by species, dimensions and grade, can be made available within given time periods at tentative mill sites within alternative cost limits per volume unit?". The answer to this question depends on the type of industry that is going to use the available stock. The requirement for a paper mill, plywood mill or sawmill are quite different in terms of acceptable species, dimensions and grades.

If the owner of a forest tract wants information aimed at management decisions, then the inventory should be planned to obtain estimates of the volume growth and the quantity of timber cut. Other types of forest inventory exist and some are discussed by Husch (1971).

It must be emphasized that the information generated by alternative types of forest inventories is not interchangeable. Although one can make use of existing information, including past inventories regardless of their type, it is advised to make a specific inventory designed for each situation (Nyyssonen, 1976).

1.1 Statement of the Problem and Assumptions A forest inventory constitutes the basic step in order to obtain information on the availability of timber products in a specific area. Government agencies, private companies or consulting foresters may be involved in planning and executing forest inventories. There are several sampling methods that are available for conducting forest inventories. Depending on the objectives of the decision maker and taking into consideration factors such as size and accessibility of the area, forest types present and funds available, an appropriate selection of a sampling method can be done.

The decision maker may be faced with the following problem: selecting a multilevel sampling technique for reconnaissance inventory when a large area to be inventoried is involved and limited funding is available. In such a case, either a multistage or multiphase sampling techniques can be considered.

In order to give guidance for the decision maker in selecting the most appropriate multilevel sampling technique, a multistage and a multiphase method were chosen from the several multilevel sampling methods available to be evaluated on the present research project. The requirements for the selection of the methods were: a) they should allow for the use of remotely sensed imagery, including airphotos and satellite imagery because such images are convenient when large areas to be inventoried are involved; b) since satellite imagery is available in printed, transparent and digital format and the organization responsable for the inventory may or may not have facilities for processing digital imagery, the satellite imagery to be used on the methods should be in the form of printed or transparent images as well as digital, and c) the methods should be applicable in situations where stratification of the forest is possible.

Based on these requirements, the multistage method selected was the one developed by Langley (1975) and the multiphase method was the one developed by Johnston (1982). The first method has been used more frequently for inventories in the west coast region of the United States. The second method was only recently developed. No record of its (practical) use could be found except with simulated data (Johnston *et al.*, 1983).

The approach taken to help the decision maker in selecting between the two multilevel sampling techniques, was to independently apply them on an area located on the Manistee National Forest in northern lower Michigan. The size of the area is equivalent to two townships. Sample plots were independently selected for each method and, within each plot, two randomly chosen point samples were used to sample volume. Regional volume equations were used to estimate per acre volume. Stratification of the area was done by photointerpretation of 1:24,000 CIR (color infrared) transparencies. As the multiphase method requires the use of digital imagry, the processing of this image was performed on the ERDAS-400 microcomputer at the Center for Remote Sensing at Michigan State University. In order to estimate the cost for each method, the time required to perform the several activities on each of the procedures was measured. The evaluation of both procedures was done in terms of the precision of the estimated total volume calculated for each method and by the total cost as estimated by measuring the field and office time. In order to estimate the total cost for each method, an average price per hour charged by consulting foresters working in Michigan was used. This average price was estimated by a telephone survey conducted among registered consulting foresters in Michigan.

For the purpose of this research project, several assumptions were made. First, it was assumed that the objective was to perform a reconnaissance inventory of the area. As previously defined, this type of inventory is done when an estimation of total volume of standing trees is desired. Second, the estimated volume would refer to the pulpwood values given in cubic feet per acre. Finally, in order to estimate costs, the work would be performed as if it were done by an individual consulting forester.

The results of this research are expected to be of value to future users of the methods in inventories in the northern region of Michigan.

CHAPTER 2

LITERATURE REVIEW

2.1 Background In the planning phase of a forest inventory it is necessary to define, among other things, the unit of assessment. If a national forest inventory is to be executed, the unit of assessment could be a state or region. In smaller properties, it could be the whole property, a stand or areas with similar forest types, age or site index. These units may be delineated on a map or on any form of remotely sensed imagery (e.g. airphotos, satellite or radar imagery).

The use of remotely sensed imagery, mainly airphotos, provides an immense amount of information for the execution of a forest inventory. By means of photointerpretation, the area can be stratified into several forest types. Nonforest areas may also be identified and segregated since these areas may not be of immediate interest. Measurements of crown diameters, crown closure and, in some cases, tree height can be made precisely.

Although all these data can be obtained from airphotos, it is only in rare instances that timber volume can be directly estimated (Husch, 1971). As a result, field measurements are always part of any forest inventory.

If the forest area is small or the trees very valuable, all the trees on the property could be measured. In this case we would obtain a census or a complete enumeration. Care must be taken in such cases not to measure the same tree twice or to omit any of the trees (Husch, 1971). Normally larger areas of forest lands, involving hundreds or thousands of acres, are the object of the inventory. In these cases, a census would not be feasible and a sampling method has to be chosen in order to estimate the parameters of interest.

In applying any sampling method, it is essential that the population be clearly defined. The population is considered as the aggregate of units from which the sample is chosen (Freese, 1962; Cochran, 1977). The sampling units are the minimum size unit into which the population is divided (Chapelle, 1985). In a forest inventory the area of interest (population) is divided into sampling units with a specified size. This size has to be kept constant otherwise it is not possible to apply the laws of probabilities of sampling (Loetsch et al. 1964).

Once the population is defined, it is subdivided into sampling units of a specified size and shape. This subdivision may be based on past experience with common forest types or upon results of previous inventories under similar conditions. Independent of the size or shape of the sampling units, unbiased estimates of the parameter of interest can be obtained. What will be affected by the shape and size of the sampling units are the cost of the survey and the precision of the estimate (Husch et al. 1982).

If a map of the property to be inventoried is available, it can be used to subdivide the area into the sampling units. One restriction is that the units must cover the whole population and must not overlap (Cochran, 1977). The map showing the sampling units constitutes the frame or the list of all units on the population. It is from this frame that the sample of units will be selected for field measurements. A sampling frame can also be established by using airphotos or other forms of remotely sensed data such as satellite or radar imagery. The latter are used when large areas covering thousands of acres are to be inventoried.

There are several sampling methods that can be applied to forest inventory. The selection of the most appropriate method depends on the area of interest: whether the area is composed of only one or of several forest types; the extent of the area; the objectives of the inventory; the costs involved; the availability of remotely sensed imagery.

The sampling units within each sampling method may be selected randomly or systematically. In random selection, the laws of probability are involved whereas in systematic selection, once the first unit is selected the positions of the remaining units are automatically defined.

2.1.1 Simple Random Sampling (SRS) - This method consists of selecting sampling units from the population such that every possible combination of n units has an equal chance of being selected (Freese, 1962; Cochran, 1977). The advantage of the SRS method is that it yields an unbiased estimate of the parameter of interest and allows the estimation of the sampling error, which is a measure of the precision of the estimate.

The sampling units can be selected with or without replacement. In the first case, a unit can appear in the sample more than once and in the second, a selected unit may appear in the sample only once. The formula for computing standard error is different depending upon which replacement system is used; the formula is simpler for sampling with replacement (Freese, 1962; Cochran, 1977).

The use of SRS has some disadvantages such as the traveling time between dispersed sampling units and the possibility of selecting sampling units which may result in atypical estimates of the parameter of interest when the sampling units are concentrated in areas of high or low volume (Husch et al. 1982).

In applying SRS for selecting units in a forest inventory, some of the sampling units may have different sizes. This happens when the areas of interest have irregular shapes and some of the measurement units fall on the border of the area or, in case of strips, their total length does not fall within the limits of the area. In this case, the laws of probability can not be rigorously applied since they require that all sampling units be of the same size (Loetsch *et al.* 1964).

To solve this problem, ratio or regression estimators may be used. In ratio estimation, two related variables have to be enumerated on the same unit, such as area of the unit and the corresponding timber volume. The ratio estimator in this case would be the ratio of the mean volume to the mean area per sampling unit. In order to estimate the total volume of timber, the total area of the forest must be known. The formula for the estimated variance of the total depends on the precision of the area measurement. If the area is measured without error, the formula is simpler than if the area is an estimate from a sample and therefore subject to sampling error. Husch *et al.* (1982) and Loetsch *et al.* (1964), present such equations.

The ratio estimator is biased although the bias becomes negligible as the sample size increases (Cochran, 1977). In order to use the ratio estimator effectively it is required that a linear correlation exists between the two variables and that the regression line goes through the origin (Freese, 1962; Loetsch *et al.* 1964; Cochran, 1977; Husch *et al.* 1982). If this is not the case, a regression rather than a ratio estimator should be used.

As with a ratio estimate, the regression estimate is used to increase precision by the use of an auxiliary variable x_i that is correlated with y_i (Cochran, 1977). If, for instance, x_i is the area of the sampling unit and y_i is the corresponding volume, the estimated regression coefficient would express the average change in volume per unit change in area between the sampling units in the sample and the population (Husch *et al*. 1982).

There are certain conditions that must be met in order to use the regression estimator (Freese, 1962). The population mean for the independent (supplementary) variable must be known; the relationship between the dependent and the independent variables has to be reasonably linear; and the variance of the dependent variable about its mean should be constant for all values of the independent variable.

2.1.2 Stratified Random Sampling - This method is applied when the forest to be inventoried can be subdivided into smaller areas of more homogeneous characteristics called strata. An independent random sample is drawn within each stratum.

According to Freese (1962), stratified random sampling presents several advantages over simple random sampling. It allows separate estimates of the mean and variance for each stratum and, for a given sampling intensity, it gives more precise estimates of the parameter of interest.

Some requirements have to be met in order to obtain a more efficient sampling by means of stratification. These are (Bickford, 1961): 1) the strata must be define independently from the sampling; 2) there must be real differences among strata means and variances; and 3) there must be a proper distribution of units per stratum.

Several criteria may be used to separate the area into homogeneous strata, such as topographical features, forest types, density classes, height, age, site classes, geography, and stand conditions, (Bickford, 1961; Husch *et al*. 1982).

Stratification is normally done by photointerpretation. In fact, this is one of the most valuable uses of airphotos for forest inventory. When possible, the criterion used to stratify should be closely related to the parameter of interest, since this would give large gains in precision (Bickford, 1961; Cochran, 1977; Husch *et al*. 1982).

When a heterogeneous population is stratified the total variation is divided into two parts: the variation within strata and the variation between the strata (Loetsch *et al*. 1964). If the variation among units within a stratum is less than the variation among units from different strata, the population estimate will be more precise than if simple random sampling is used (Freese, 1962). As the variation within a stratum is reduced, since it is internally homogeneous, a precise estimate of the stratum mean can be obtained from a small sample (Cochran, 1977). The greater the difference among stratum means, the more advantage there is to use stratified sampling in comparison to simple random sampling (Bickford, 1961).

The third requirement relates to the distribution of sample units within each stratum. Several procedures may be used to allocate the total number of sampling units to each stratum such as: proportional, optimum, and balanced allocations (Freese, 1962; Husch et al. 1977).

In proportional allocation, the number of units in a stratum is proportional to the relative area of the stratum (Husch et al. 1982). This allocation procedure requires advanced knowledge of the variance within each stratum in order to estimate the total number of samples to measure. If this is not available, the total sample size can be calculated as if a simple random sampling was going to be used for the inventory (Husch et al. 1982).

Optimum allocation is used when costs are involved, or when we want to obtain a precise estimate of the stratified mean for a fixed cost (Husch et al. 1982). In optimum allocation, the distribution of sampling units per stratum depends on whether the cost per sampling unit varies or is the same in each stratum. In the first case, the sample size per stratum is inversely proportional to the cost per sampling unit. In the second case, the sample size is proportional to the product of the stratum area and its standard deviation. This last case is also called Neyman allocation (Cochran, 1977).

In balanced allocation an equal number of sampling units is taken on each stratum. This type of allocation is less efficient than proportional or optimum (Bickford, 1961).

2.1.3 Cluster Sampling - In some cases of forest inventory it may be costly or impractical to select a sample by the SRS method. This may happen when the distance between sampled units is so large that the time and cost of covering the entire area is prohibitive, or when it is impossible to identify each unit in the universe (Frayer, 1979). In such situations, the sampling units can be aggregated into a number of mutually exclusive groups or clusters of such units. Each cluster should have an equal number of units. A simple random sampling of each group is drawn and all the units within each selected cluster are measured. This process is called one stage or simple cluster design (Frayer, 1979).

When the units of assessment are large, such as in regional or national forest inventories, the measurement of all units within a cluster is costly or impracticable. It is convenient in such cases to subdivide the clusters into units of hierarchical order. This process constitutes what may be called multilevel sampling designs.

2.1.4 Multilevel Sampling Designs - The term multi-level implies that more than one source of information will be used in the estimation of population parameters (Frayer, 1979). These sources generally, but not necessarily, involve one or more types of remotely sensed imagery (e.g. satellite images, high, medium and low altitude aerial photographs) and ground measurements.

Multilevel sampling may be classified into multistage sampling and multiphase sampling. In multistage sampling designs, a population is divided into a number of primary sampling units, each of which is subdivided into smaller secondary units. The secondary units may in turn be subdivided into smaller, tertiary sampling units (Husch *et al.* 1982). At each stage, an independent sample is taken by random or systematic selection (Loetsch *et al.* 1964). The random selection process can be done with equal or varying selection probabilities. The last is also called sampling with probability proportional to size (PPS).

In multiphase sampling, information on auxiliary variables is used from various phases in estimating population parameters (Frayer, 1979, 1981). The difference between multistage and multiphase sampling is in the size of the "sampling unit". For multiphase sampling unit size remains the same, independent of the number of phases of information used. A first phase unit is the same size as a second phase unit, and so on. In multistage sampling, the units are partitioned into smaller units at each succeeding stage.

There are two types of multiphase sampling, those using regression estimators and those using stratification. Regression estimators are used when the auxiliary variable is, for instance, a photo plot estimate of the variable of interest (i.e., volume or density). Stratified estimators are used when the auxiliary variable is an indicator random variable, specifying into which stratum a sampling falls at a given phase (Johnston, 1982).

2.1.5 Double Sampling - This is a form of multiphase sampling with only two phases. This method is suitable for the use of regression estimators when the population mean and total of the auxiliary variable is unknown (Freese, 1962).

In the first phase, a large random sample is taken from the auxiliary variable (independent variable) in order to have a precise estimation of its population mean and total. In the second phase, a random subsample is taken from the first sample and the variable of interest (dependent variable) is measured (Husch *et al*. 1982).

A regression equation can then be developed using the measurement on both variables. The form of the regression equation does not necessarily have to be linear.

The use of remotely sensed imagery such as airphotos is possible with this sampling method. The auxiliary variable could be, for instance, the estimated volume per unit area obtained from photo interpretation of photo plots (Husch *et al*. 1982).

2.1.6 Systematic Sampling - Systematic sampling is a form of nonrandom sampling. The units included in a systematic sample are not selected at random but according to a pre-specified pattern (Freese, 1962). The population, numbered 1 to N, is split into n groups of K units each where K=N/n. A random number is selected between 1 and K, say r^*, and the r^{th} element of each of the K groups comprise the sample (Cochran, 1977).

Systematic sampling has the advantages of providing unbiased estimates of homogeneous population means and totals, and being faster and cheaper to execute compared to other probability sampling methods (Husch et al. 1982). Additionally, travel between sampling units is usually faster because their distribution pattern facilitates their locaiton. Lastly, the size of the population need not be known since they are selected at fixed intervals.

The great disadvantage of systematic sampling is that there is no valid method for estimating sampling error (Husch et al. 1982; Cochran, 1977; Frayer, 1979). The reason for this is that the calculation of variance requires a minimum of two randomly selected sampling units. In systematic sampling, only the first unit is randomly selected, the others being taken at a constant interval.

Another disadvantage is that systematic sampling may give poor precision when unsuspected periodicity is present in the population (Cochran, 1977). In this case, the variance of the estimators change depending on the correspondence between the periodic nature of the population and the sampling interval (Frayer, 1979).

If the population of interest has a linear trend, a systematic sample is more effective than a simple random sample (Cochran, 1977; Frayer, 1979). If the population is in "random" order, then systematic sampling would be equivalent to simple random sampling. This is one condition where an unbiased estimate for the variance of the estimator exists and can be calculated by using the equations derived for simple random sampling.

2.2 Multistage Sampling Technique Multistage sampling may be used for surveying vegetation on large areas about which little is known. The population is divided into large primary sample units which are selectively subdivided into smaller and smaller units. Each subdivision constitutes a stage (Anderson, 1979; Meyers et al. 1980; Husch et al. 1982).

The multistage sampling technique has been used in forestry for many years. Freese (1962) and Loetsch et al. (1964), discussed the applications of two and three stages sampling techniques to forest inventory. The primary units were composed of blocks of equal or unequal sizes, which can be delineated on maps of the area or on airphotos.

The use of airphotos in multistage sampling for forest surveys may or may not be necessary, depending on the development of the method. Multistage sampling techniques have been developed that do not require airphotos for their application. Such development was done by Frayer (1979) and Johnston (1980). They described multistage sampling techniques, up to four stages, for resource inventory. Formulas were derived for the estimation of means, totals and proportions, and the respective variances. The application of the methods developed was illustrated with simulated numerical examples.

Yandle *et al.* (1977) developed a simpler two stage sampling procedure for forest inventory which does not make use of aerial photography. The first stage is point sampling using a relascope, the second stage involves list sampling with replacement and with selection probabilities proportional to size. This method is more appropriate for updating a survey or in surveys in which the volume is estimated by measuring a subsample of trees from the first stage.

If aerial photography of the area to be inventoried can be obtained at several scales, each scale may be used as a stage. The number of stages one can include is not proportional to the size of the area to be inventoried. Two and three stages sampling designs using airphotos and ground measurements have been used in forest surveys of large areas, varying between 100,000 to 1.6 million acres. The objectives of the surveys varied but satisfactory results in terms of precision of the estimated parameter of interest and cost were obtained (Wert, 1968 and 1969; Heller *et al.* 1969; Hall *et al.* 1979; Harris *et al.* 1983).

The launch of spaceships such as in the APOLLO program, the SKYLAB and more recently the LANDSAT series of satellites, opened a new perspective for the application and development of multistage sampling techniques. The images obtained from space are normally used as a first stage in multistage designs. Although the resolution of such images may not be appropriate for detailed photointerpretation, the advantage that they offer is in terms of the large area covered. One image may contain the entire area to be inventoried. The introduction of subsequent stages with increased image resolution, allowing for more detailed photointerpretation, may increase the efficiency of sample selection at each stage.

Langley (1969, 1971 and 1975) developed a multistage variable probability sampling technique which may incorporate satellite images or high-altitude photography as a first stage. His method proved to be appropriate for inventorying large areas and has been successfully used in several occasions to estimate the total volume of an area (Wert, 1968 and 1969; Anderson, 1979; Lee et al. 1984).

Satellite images can be obtained as a printed image or as a digital image. Both forms of the image can be used as a first stage in multistage sampling techniques. Nichols et al. (1973) used automatic classification of LANDSAT digital data as a first stage in a three stage sampling design. The classified image was subdivided into four timber volume classes. The second stage was composed of aerial photography of the selected primary units. The third stage was the ground measurements. The method was shown to be cost effective when compared to the 10 points system applied to the same area. The sampling error was 8.2%, which was far below the acceptable value of 20% for the area inventoried.

Multistage sampling designs have been applied not only to estimate total volume but other parameters of interest as well. Hall *et al.* (1979) used a two stage sampling procedure for collecting data on vegetation parameters such as height and width, percent cover, relative density and frequency by species for general land use planning with emphasis on grazing and wildlife management. Color infrared photographs (CIR) at a scale of 1:12,000 and line transects were used as the first and second stages, respectively. Although the cost per unit area of the inventory was considered high, the CIR photographs proved useful for inventory stratification.

Gialdini *et al.* (1975) described the steps to be followed in the design and implementation of an imagery-based information system. The author described three case studies in order to demonstrate the applicability of the system for timber management planning. The first case study discussed was the inventory described by Nichols *et al.* (1973), where the sampling procedure used was appropriate and the LANDSAT imagery proved useful as a first stage. Estimating only the total merchantable volume, the objective of this particular study, might have not been enough for timber management purposes. On the second study, other parameters such as number of acres, number of trees, basal area, basal area growth and surface area were estimated in addition to volume. In this case, a five stage stratified sampling procedure was used. LANDSAT data were used on the first stage but were not used to calculate selection probabilities. Instead, LANDSAT data were used for stratification and for area measurement within strata, based on computer classification. The reason for this change in the use of the LANDSAT data was that in multiparameter surveys, inflated variances can be obtained for some parameter if probability proportional to size sampling is used and if the variables are not all positively and highly correlated with each other. The overall results of the survey were considered satisfactory, although the relative standard error associated with some parameters was above the 5% specified for the volume. Langley (1978), discussed the possibilities of using remote sensing in multistage sampling methods for multiresource inventories, and presented the difficulties involved in such inventories.

The third case study was discussed in more detail by Titus et al. (1973). For this analysis, SKYLAB S190 CIR photography was used instead of LANDSAT digital data as the first stage in a three stage sampling method. Other objectives evaluated in the analysis were: 1) the efficiency of manual photointerpretation of enlarged S190 CIR imagery in identifying timber volume classes and 2) the application of simple random sampling for selecting sampling units at each stage, as opposed to PPS (probability proportional to size) sampling. The results showed that the correlation between the ground condition and the photointerpreter's timber volume class estimations were very low. A division into two classes (timber and non-timber) seemed more promising for a first stage stratification, due to the higher correlations with ground conditions. This suggested that PPS sampling would be more efficient than simple random sampling. The probability of selection would be proportional to the percent of the primary unit covered by timber.

The feasibility of using LANDSAT data in a multistage sampling design for regional forest inventory was discussed by Nichols et al. (1976) and Harding et al. (1978). The area involved was 10 million acres located in western Washington. The objectives of the inventory were to provide a forest volume data base for a large region, by broad ownership class, within a limited time and to an acceptable level of accuracy. Basal area per acre was estimated by five ownership categories for both second growth conifer and hardwood. The specified precision of 10% for the estimated mean was achieved for four of the five ownership categories. The study did not positively confirm the feasibility of using LANDSAT data for forest management inventory but it showed the promise of such data. Further research was to be carried out to compare a detailed LANDSAT classification with the current forest management inventory.

Stratification is a common practice in forest inventory. The area to be inventoried is subdivided into strata of homogeneous composition, based on specific criterion. This allows for a reduction on the total number of sampling units, since they will be located on areas, or strata, of more homogeneous variance. The end result is a reduction on the cost of the survey. Multistage sampling has been used with satisfactory results in situations where the area to be inventoried was stratified (Wert, 1968; Heller et al. 1969; Titus et al. 1973; Gialdini et al. 1975; Langley, 1975; Harris et al. 1983). On other applications, multistage sampling techniques have been combined with multiphase techniques with satisfactory results (Hegyi, 1980; Peterson et al. 1983).

Although multistage sampling has been successfully used for forest surveys for different objectives, the technique has some disadvantages. One disadvantage is the localized distribution of ground plots. Because of this aspect, the use of multistage technique may not detect as much of the variability in the population as would an equal number of plots randomly distributed over the area. As a result, the sampling error for multistage sampling tends to be larger than that for a completely random sampling with the same number of ground plots (Meyers et al. 1980). Another disadvantage of multistage sampling is related to its application when the population is stratified. If stratification is done by vegetation type, it tends to interfere with the efficiency of the method (Aldred, 1980).

The decision to apply a multistage procedure to forest inventory is basically an economic one. In some situations the parameter of interest can be estimated within the established precision and cost by using only one stage. However, as the area to be inventoried gets larger, the introduction of more stages becomes more feasible (Langley, 1978). On the other hand, since the ratio of travel costs to the costs for installing a field plot is high for multistage design, the method, despite its disadvantages, tends to be used (Meyers et al. 1980).

2.3 Multiphase Sampling Technique Multiphase sampling makes use of information in auxiliary variables from various phases, or levels, in estimating population parameters (Frayer, 1979, 1981).

If an n-phase sampling design is used, at each of the first n-1 phases selected, sampling units are located on the corresponding levels of imagery. An auxiliary variable, correlated with the variable of interest, is measured. At the final phase, ground measurements may be taken (Johnston, 1982).

The difference between multistage and multiphase is in the size of the "sampling units". In a multiphase design the unit size is constant, it is independent of the number of phases of information used. A first phase unit is the same size as a second phase and so on. In multistage sampling, the units are partitioned into smaller units at each succeeding stage (Frayer, 1979, 1981).

There are two types of multiphase sampling designs: those using regression estimators, also called with dependent phases and those using stratification, also called with independent phases (Loetsch *et al.* 1964). The first type is used when the auxiliary variable is a photo plot estimate of the variable of interest (i.e. volume or density), which is measured on the ground. The second type is used when the auxiliary variable is an indicator variable, indicating into which stratum a sampling unit falls at a given phase (Johnston, 1982). Most of the literature on the subject deals with the first type of multiphase sampling, while few articles discuss the second type (Johnston, 1982).

Double sampling is a form of multiphase sampling limited to only two phases. Regression or ratio estimators can also be used with double sampling (Hush *et al.* 1982). Regression estimators are quite useful when aerial photographs of the inventory area are available. In such cases, a large number of photo plots are randomly selected as the first phase. A supplementary or auxiliary variable is then measured. For example, plot volume is estimated by photointerpretation. For the second phase, a subsample of the photo plots is selected for field measurement of the variable of interest. A linear regression equation is then developed between data from the plots measured on the first and second phases. The equation can be utilized to estimate the mean and the total of the parameter of interest. The relationship between the two variables need not be linear. This type of sampling design is useful when the measurement of the variable of interest is costly or for updating a forest inventory (Freeze, 1962; Hush *et al*. 1982). The location of the photo plots on the first phase can be done systematically rather than randomly, with multiple random starts to overcome the problem of estimating the sampling error (Shine *et al*. 1962).

Double sampling is a common sampling method. It has been shown to be very efficient in terms of cost and precision of the estimate when applied in either temperate or tropical forests (Bickford *et al*. 1963; Temu *et al*. 1971; Hutchinson, 1978; Mattila, 1984). Double sampling can be used when the inventory unit has been stratified (McLean, 1972; Cochran, 1977; Johnston, 1982). In some applications, double sampling with regression estimators has been shown to be more efficient in terms of cost and estimated sampling error than simple random sampling (Wear *et al*. 1964).

Multiphase sampling is not limited to two phases. Three and four phase methods have also been developed and applied in practice (Loetsch *et al*. 1964; Frayer, 1979; Johnston, 1982; La Bau *et al*. 1983; Jeyaratnam *et al*. 1984; Li *et al*. 1984).

As seen from the literature review presented, multi-level sampling techniques have many applications in the inventory of natural resources. The introduction of satellite imagery has greatly contributed to the development of new multilevel designs. These images have the advantages of covering large areas and of allowing the use of computer assisted classification, which is used as a first level on several of the methods discussed. With the development of satellite sensors with higher resolution, such as SPOT (Systeme Pour l'Observation de la Terre) with a resolution of 10x10 meters in one of its sensors (Jensen, 1986), the application of multilevel sampling techniques may become more efficient due to the better capability for stratification either visually or automatically.

It is possible to combine the available sampling methods with both types of multilevel designs, thereby creating more complex sampling techniques. If a simpler sampling procedure will achieve the inventory objectives, then it should be used in place of a more complex design. The selection of an appropriate sampling technique, be it a multistage or a multiphase design, is facilitated if previous successful experience with the method on similar conditions is available. Some methods have only been used on simulated conditions and, as such, their practical usefulness cannot be properly evaluated. Other methods, although widely used, may not have been applied on certain regions. As a result, their applicability on such areas is convenient to evaluate.

CHAPTER 3

METHODS

This chapter presents the general aspects of the sampling methods selected for evaluation. In the first section, the multistage sampling technique is presented. This technique was developed by Langley (1975). It is a variable probability method, i.e. the probabilities of the sampling units being selected at each stage varies. Three stages were considered in the present study. A sensitivity analysis was performed on each sampling technique evaluated and is also described in the first section.

The second section describes the multiphase sampling technique selected. This technique was developed by Johnston (1982). It is a double sampling for stratification and may be considered as an extension of stratified random sampling, wherein a two phase sample rather than a simple random sample is taken in each stratum.

The third section gives a description of the area where the selected sampling techniques were independently applied. This area is located in northern Michigan. The fourth section describes the remote sensing imagery (air photos and satellite imagery) used for the project. It also presents the method used for field measurements.

The fifth section describes the methods used for photointerpretation and digital processing of the imagry of the study area. This section also includes the following parts: description of the process used for sampling selection at each stage of the multistage sampling technique and at each phase of the multiphase sampling technique, the equations used for volume estimation, the process for area estimation and time and costs considerations.

3.1 Multistage Sampling Technique In multistage sampling, a sample of first stage units or primary sampling units (PSU) is randomly selected. Within each PSU selected, a sample of secondary sampling units (SSU) is also randomly selected. The process continues until sampling units have been selected from each stage. The variable of interest is then measured in the last stage units (Johnston, 1982).

The above description of multistage sampling technique refers to randomly selected samples at each stage. The process of random selection implies an equal probability of selecting every possible combination of n sampling units from the population (Freese, 1962; Hush et al. 1982). If selection is made with replacement, or the same unit can be selected more than once, the constant probability of selection is given by P=1/N, where N is the total number of units in the population (Anderson, 1979). The estimate of the total timber volume when the probability of selection is constant can be expressed as (Langley, 1975; Anderson, 1979):

$$y = (1/n) \sum_{i=1}^{n} (y_i/P) \qquad (1)$$

where y_i is the estimated total timber volume of the i^{th} unit;

n is the number of sample units measured;

P is the constant probability of selection.

The term y_i/P is an estimate of the population timber volume Y, based on the measurements made on each of the n sampling units (Anderson, 1979).

Another alternative is to select sample units by variable probability. This process is called selection with probability proportional to size - PPS (Cochran, 1977). In this case, the probability of selecting the i^{th} sample unit at the j^{th} draw is derived from a criterion such as predicted timber volume obtained from aerial photographs (Langley, 1975). This probability is given by:

$$P_i = x_i / \sum^{N} x_i \qquad (2)$$

where x_i is the predicted volume or some other characteristic of interest and ΣP_i ($i = 1,\ldots,N$) is equal to unity.

By substituting equation 2 into equation 1 the new estimator of the total timber volume will be given by:

$$y = (1/n) \sum_{i=1}^{n} y_i/P_i \qquad (3)$$

The estimator y is unbiased if the sample units are drawn with probabilities P_i and with replacement (Cochran, 1977). Equation 3 reflects the simplest case of multistage sampling where only one stage is considered.

Any estimate of a population parameter has an associated measure of variability. In the case of one stage PPS the variance of the estimate given by equation 3 is (Langley, 1975; Cochran, 1977):

$$\mathrm{var}(y) = (1/n) \sum_{i=1}^{N} P_i [(y_i/P_i)\ y]^2 \qquad (4)$$

where N is the number of units in the population and the other terms are as already defined.

As seen from equation 4, the variance of y depends on the squared deviations of the ratios y_i/P_i from y. If the determination of the probabilities of selection of the i^{th} unit is based on a variable with a high correlation with the variable of interest, the less variability there is among the ratios y_i/P_i. As this ratio is an estimate of the parameter y, the squared deviations would be small and, as a result, the variance of the estimator would be reduced (Langley, 1975; Anderson, 1979).

If the area to be inventoried is large, resulting in large primary sampling units, a second or third stage may be necessary. In the present study, three stages where used in order to estimate the total volume. The estimator and its respective variance are given as:

$$y = 1/m \left(\sum_{i=1}^{m} 1/P_i * n_i\right)\left(\sum_{j=1}^{n_i} 1/P_{ij} * t_{ij}\right)\left(\sum_{k=1}^{t_{ij}} y_{ijk}/P_{ijk}\right) \quad (5)$$

$$\begin{aligned} var(y) = {} & 1/m \sum_{i=1}^{M} P_i[(y_i/P_i) - y]^2 + \\ & 1/m \sum_{i=1}^{M} 1/P_i n_i \sum_{j=1}^{N_i} P_{ij}[(y_{ij}/P_{ij}) - y_i]^2 + \\ & 1/m \sum_{i=1}^{M} 1/P_i n_i \sum_{j=1}^{N_i} 1/P_{ij} t_{ij} \sum_{k=1}^{T_{ij}} P_{ijk}[(y_{ijk}/P_{ijk}) - y_{ij}]^2 \end{aligned} \quad (6)$$

where: m = number of primary units on the sample

n_i = number of secondary units drawn within the i^{th} primary unit

t_{ij} = number of tertiary units drawn within the j^{th} secondary unit within the i^{th} primary unit

y_{ij} = volume measured on the j^{th} secondary unit of the i^{th} primary unit

P_{ij} = selection probability of drawing the j^{th} secondary unit given by the i^{th} primary unit

y_{ijk} = volume measured in the k^{th} third stage unit in the j^{th} second stage unit in the i^{th} primary unit

P_{ijk} = probability of selection associated with y_{ijk}.

The other terms are as previously defined. The probability P_{ij} and P_{ijk} are given respectively by:

$$P_{ij} = x_{ij} / \sum_{j=1}^{N_i} x_{ij} \qquad (7)$$

$$P_{ijk} = x_{ijk} / \sum_{k=1}^{T_{ij}} x_{ijk} \qquad (8)$$

where: x_{ij} = the characteristic of interest for the j^{th} second stage unit within the i^{th} first stage unit

x_{ijk} = the characteristic of interest for the k^{th} third stage unit within the j^{th} second stage unit within the i^{th} first stage unit. The formulas to be used to calculate the optimum allocation of sampling units to each stage are presented on Appendix A.

3.1.1 Sensitivity Analysis A sensitivity analysis was performed in order to evaluate what effect a change in the selection probabilities for the first (P_i), second (P_{ij}) and third stages (P_{ijk}) would have on the estimated total volume and its variance. This was done by randomly changing the values of the size variables for each stage. A table of random numbers was used to derive the following rule: if the random number is even, increase the value of the size variable by 10%; if the random number is odd, subtract 10%. The reason for using 10% was to be compatible with the intervals on the crown density scale used.

After applying this rule for each stage within each stratum, the probabilities were recalculated and entered in equations 5 and 6 for the estimation of the total volume and its variance, respectively. A total of five trials were performed on each stage within each stratum, except for the first stage of the red oak stratum. Since only two first stage units (primary units) were contained within this stratum, only three trials could be obtained that did not repeat the same value of the selection probabilities. When the selection probabilities for a given stage were being changed to calculate the total volume and its variance, the probabilities for the other stages were kept with their original values.

In order to further evaluate the effect of changing the probabilities on the estimated total volume and its variance, those probabilities were combined among themselves. A total of 125 combinations of the probabilities P_i, P_{ij} and P_{ijk} resulted from the five trials for the conifer and hardwood strata. Only 75 combinations resulted from the red oak stratum, since only three trials were possible for the first stage, as explained above.

The 125 combinations for the conifer and hardwood strata formed five groups, each with 25 combinations of the selection probabilities. In the combinations of the probabilities P_i, P_{ij} and P_{ijk} for group A, the probabilities P_i were kept constant and equal to the values for trial# 1 shown in Appendix B.

The following example will illustrate the way the combinations were performed. Within group A, the following combination among P_i, P_{ij} and P_{ijk} was obtained: 1, 3, 5. The number 1 refers to the two values of the probability P_i for trial #1 shown on Appendix B, for a given stratum. These values were kept constant for the other 25 combinations within group A. The number 3 refers to the four values of the probability P_{ij} for trial #3. The number 5, refers to the twelve values of the probability Pijk, for trial #5.

The set of 25 combinations for each of the other trials B, C, D and E were obtained in the same way except that the probability P_i would refer respectively to the values calculated for trial #2, #3, #4 and #5.

The three trials for the red oak stratum were called A, B and C and were also formed by a set of 25 combinations of the probabilities P_i, P_{ij} and P_{ijk} in the same way as explained above. For each of the 325 total combinations for the three strata, the total volume and its variance were calculated.

3.2 Multiphase Sampling Technique The multiphase sampling technique selected can be generally described as follows: consider a universe of N sampling units which is stratified into L population strata. Each strata contains a known number of units N_i ($i = 1,\ldots,L$) such that:

$$N = \sum_{i=1}^{L} N_i \qquad (9)$$

At the first phase, a simple random sample of size n_i is drawn within the i^{th} population stratum, given by:

$$n_i = u_i * N_i \qquad (10)$$

where u_i is a sampling fraction chosen in advance for the i^{th} phase one stratum and $0 < u_i \leq 1$.

Within the i^{th} population stratum define the I_i as being the number of phase one strata. Then in the i^{th} population stratum the n_i units are partitioned into the I_i strata such that the $(ij)^{th}$ $(j = 1,..,I_i)$ phase one stratum contains n_{ij} sampling units and:

$$n_i = \sum_{j=1}^{I_i} n_{ij} \qquad (11)$$

The unknown number of sampling units in the universe contained in the $(ij)^{th}$ phase one stratum is N_{ij} and:

$$N_i = \sum_{i=1}^{I_i} N_{ij} \qquad (12)$$

At phase two, a subsample of size m_{ij} is chosen within the $(ij)^{th}$ phase one stratum:

$$m_{ij} = v_{ij} * n_{ij} \qquad (13)$$

where v_{ij} is the phase two sampling fraction and $0 < v_{ij} \leq 1$.

The characteristic of interest y_{ijk} for $k = 1,...,m_{ij}$ is then measured.

The estimators of the population strata parameters are given as:

$$y_{i..} = \sum_{j=1}^{I_i} (n_{ij}/n_i) y_{ij.} \qquad (14)$$

and

$$y_{ij.} = 1/m_{ij} \left(\sum_{k=1}^{m_{ij}} y_{ijk}\right) \qquad (15)$$

where: $y_{i..}$ = estimated population mean for the i^{th} stratum

$y_{ij.}$ = estimated population mean for the $(ij)^{th}$ phase one stratum

n_{ij} = number of sample units on the $(ij)^{th}$ phase one stratum

n_i = number of samples drawn within the i^{th} population stratum.

The variance of $y_{i..}$ is estimated by:

$$var(y_{i..}) = \frac{(N_i-1)}{N_i}\left\{\sum_{j=1}^{I_i}\left[\frac{(n_{ij}-1)}{n_i}-\frac{(m_{ij}-1)}{N_i-1}\right]\frac{(w_{ij}\ s_{ij}^2)}{m_{ij}}\right\} +$$

$$\frac{N_i-n_i}{N_i(n_i-1)}\left[\sum_{j=1}^{I_i} w_{ij}(y_{ij.}-y_{i..})^2 \qquad (16)$$

where: $w_{ij} = n_{ij}/n_i$ and

$$s_{ij}^2 = [1/(m_{ij}-1)]\sum_{k=1}^{m_{ij}}(y_{ijk}-y_{ij.})^2 \qquad (17)$$

and the other terms are as previously defined.

The unbiased estimate of the population mean is given by:

$$y_{st} = \sum_{i=1}^{L} w_i * y_{i..} \qquad (18)$$

where $w_i = N_i/N$.

The estimated variance of the estimated population mean is:

$$\mathrm{var}(y_{st}) = \sum_{i=1}^{L} w_i^2 \mathrm{var}(y_{i..}) \qquad (19)$$

where $\mathrm{var}(y_{i..})$ is given by equation 16.

3.3 Study Area

3.3.1 Location The study area (Figure 1) is located within the Manistee National Forest in Wexford County, Michigan. It includes two townships, T.21N. R.11W. (Henderson Twp.) and T.22N. R.11W. (Boon Twp). The total area of these two townships is 47,852 acres; 32,202 acres are in government ownership and 15,650 acres is private land. The present study was conducted only on government land.

This area was selected because of its location and access, the availability of information related to volume per unit area for some of the forest types, the availability of imagery, and the fact that the area also has a well developed network of roads.

3.3.2 General Characteristics The environs of the study area consists of coarse textured end moraines. Hilly topography is typical of this subdistrict of the state. Elevations range between 840 and 1,700 feet above sea level. Most of the area, both hills and depressions, is underlain by thick deposits of sandy drift and is well drained. Areas of gently sloping ground moraine and outwash plains are also present (Albert *et al*. 1976).

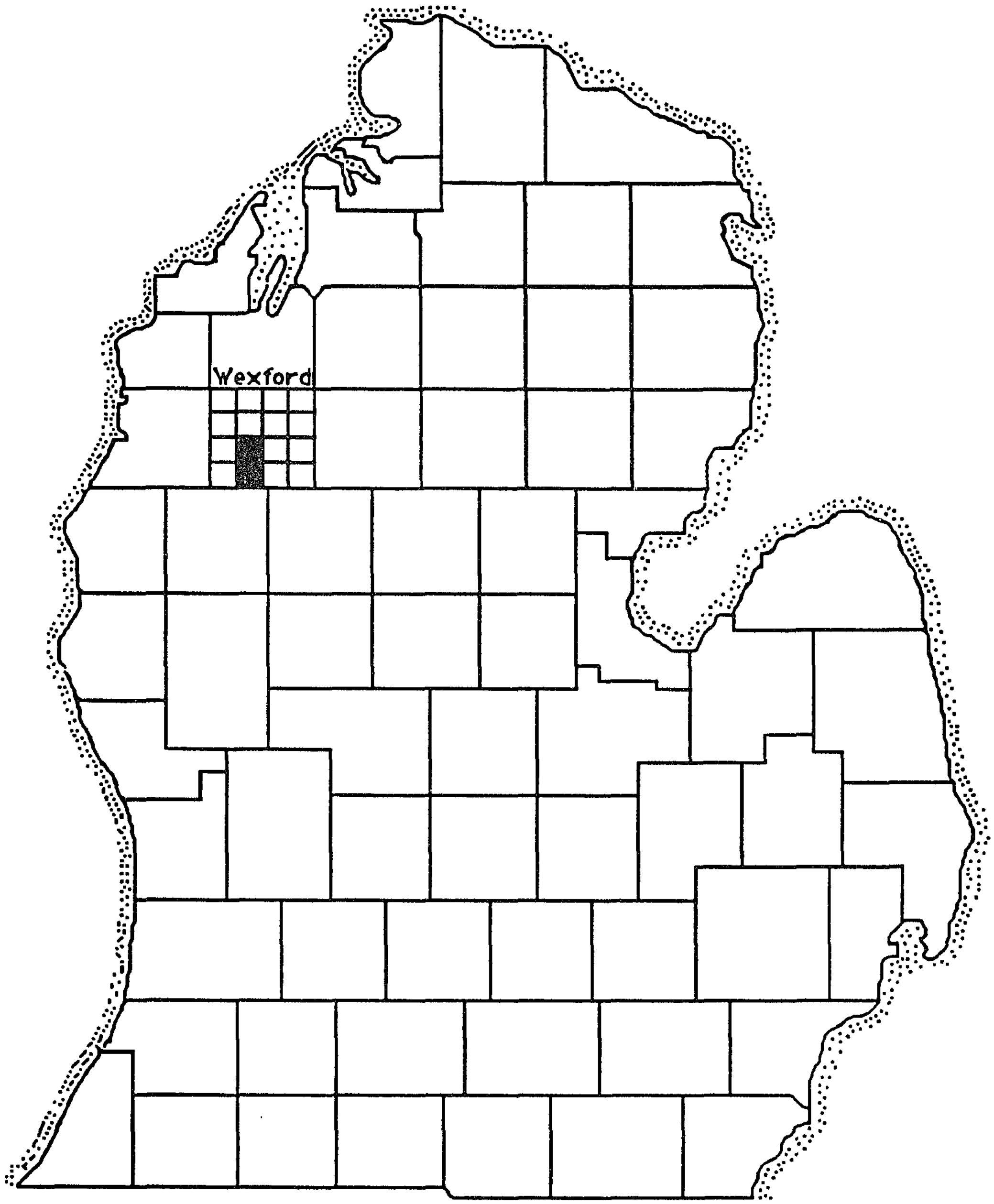


Figure 1 Location of the Study Area

The total annual precipitation in the region is 30 inches. The mean annual temperature in this part of Michigan is 44 degrees Fahrenheit, with an annual extreme minimum temperature of -18 degrees Fahrenheit.

In presettlement time, pine and pine-oak forests covered the outwash plains (see Appendix B for the scientific names) and the moraines supported beech-sugar maple forests. Logging fire and agriculture, although not intense, changed the forest composition. Early successional species such as paper birch, bigtooth and trembling aspen and red maple have greatly increased in abundance. Hemlock has been reduced due to a combination of logging, post-logging fire, and deer browsing. The occurrence of white and red pine forests, and of white pine within northern hardwoods communities has also been reduced. The conifer stands that are present in the area, most being red and jack pine, are a result of plantations established in the 1930s and 1940s. Some of these stands, mainly red pine, are now being managed by the US Forest Service.

3.4 Data Selection

3.4.1 Remotely Sensed Imagery Both sampling procedures evaluated in this study require the use of remotely sensed imagery, as already stated. For the multistage method the following imagry was used: 1) color infrared (CIR) transparencies at a scale of 1:24,000 taken in April, 1977; and 2) LANDSAT TM transparencies of true and false color composites (scene # E-40094 - 15554; November, 4 1982), at a scale of 1:750,000. For the multiphase method, the same CIR air photos were used, but the satellite data were from LANDSAT MSS in digital form (scene # 821973 - 1544 3X0; October, 30, 1981) in its four bands.

The CIR air photos were obtained on loan from the Michigan Department of Natural Resources. The LANDSAT imagery, both analog and digital, was available at the Center for Remote Sensing at Michigan State University. The late fall satellite images were used in order to discriminate the red oak stand that is located on the study area from the surrounding northern hardwood stands for purpose of stratification. The oaks have a tendency to keep their leaves longer than other northern hardwood species (Syan-Wittgenstein, 1961), a characteristic that improves discrimination during image processing.

3.4.2 Field Measurements In order to estimate the volume per unit area for each forest type, point sampling was used within the selected plots instead of measuring the whole plot for both sampling methods. This method was chosen because the field measurements would be done by only one person and because of time and financial constraints. Although point sampling might be considered as another stage in the multistage method, it is important to note that the data obtained from the sampling points can be processed in the same way as sample plot data. The estimated volume per unit area obtained by point sampling is an unbiased estimate of the population parameter of interest (De Vries, 1986).

The plot dimensions were of 73x73 meters square and 56x56 meters square for the multistage and the multiphase sampling methods, respectively.

Within each plot, two sample points were randomly located. After locating one of the corners of the plot in the field, two random numbers were selected from a computer generated list. These numbers were used as the (X,Y) coordinates (on a theoretical grid oriented to the cardinal directions) of the first point which was reached by pacing. For the selection of the second point, the same procedure was followed, by returning to the initial, marked corner. Once a pair of random numbers were used to locate the coordinates of a point sample, it was discarded.

A relascope of the Bitterlich design, incorporating a basal area factor of 10, was used for measuring all points. For each "in" tree, diameter at breast high (DBH) was always measured. But depending on the species, the height, up to 4 inches outside-bark-diameter level, was also measured with the relascope. The reason for this distinction on the variables measured was related to the type of volume equation available. For several important timber species (red pine; jack pine; black, white and red oak; aspen; sugar and red maples; cherry; paper birch and beech), local volume equations were available. For the other species a general volume equation, with height and DBH as independent variables, was used. These equations are presented in section 3.5.4.

For the estimation of the volume per unit area, only trees with DBH above 4.0 inches were considered merchantable. Appendix C presents the tally sheet used during the field work.

3.5 Data Processing

3.5.1 Photointerpretation A total of thirty 1:24,000 scale photographs covered the whole study area. Interpretation was done using a mirror stereoscope. The area was stratified into four strata: three strata were forested areas covered with mixed hardwoods, conifers and red oak; one stratum was considered as non-forest and was composed of agricultural areas and water. These strata were used in both sampling procedures. The towns located within the area were not considered for stratification purposes since they were of no interest for the project.

As only government property was included on the study, the limits of private properties from U.S. Forest Service maps were transferred to the photo-overlays. These maps, one for each township, although published in 1977, were validated by comparing them with a plat book which was published in 1983. Transfering the limits of the private properties to the overlays was accomplished using a reflecting projector at the Center for Remote Sensing. The scale of the airphotos (1:24,000) was adjusted to that of the maps (1:31,680).

3.5.2 Multistage Sampling Technique

3.5.2.1 First stage Sample Selection The first stage used the LANDSAT TM imagery in both true and false color transparencies. The images were enlarged from 1:750,000 to 1:107,000 by means of an optical projector at the Center for Remote Sensing. An overlay was placed on the rear-projection screen of the equipment and the limits of the study area were delineated. The study area is not a perfect rectangle - its northern border is approximately 484 feet shorter than its southern border. At the scale of 1:107,000, this difference corresponds to approximately 1.4 millimeters. Since the corners of the area were not well defined in the imagery, the area was considered as a rectangle with dimensions of 9.6 by 19.2 centimeters.

The area was subdivided into eighteen primary units of 2,530.3 acres each (corresponding to a square 3.2 cm on a side) which established a sample frame. An alternative size for the primary units, 1,384 acres in area, was also evaluated but rejected. Several of these smaller units would have fallen completely within private property.

Within each stratum an independent selection of two primary units was done. Time and financial constraints did not permit the selection of more primary units, which would have increased the number of tertiary units to be measured in the field.

For the calculations of the selection probabilities for the primary units, the size variable used for the three strata was the estimated total area of the forest type contained within each unit. This measurement was done with a dot grid overlay. The assumption in this case was that the more area occupied by a certain forest type within a primary unit, the greater the wood volume on that unit. Langley (1969), Titus et al. (1975) and Anderson (1979) used this same size variable for the determination of the selection probabilities. Lee et al. (1984) used percentage of reforested area per primary unit to calculate selection probabilities. This is basically the same thing, the difference being that the total area of the forest type was divided by the area of the primary unit.

The process of primary unit selection for this study followed the procedure of Langley (1975) and Anderson (1979). A cumulative list of the estimated total area for a certain forest type within all primary units was compiled. Selection probabilities were calculated by dividing the estimated total area for each primary unit by its sum. The sum of these probabilities is equal to one which satisfies the requirements of a probability distribution (Langley, 1975).

From a table of random numbers, two numbers between one and the sum of the total estimated area per primary unit were selected with replacement. A sampling unit was selected if the random number fell within its range on the cumulative list.

The false color composite imagery was used to estimate the area occupied by each forest type within the primary units. Conifers were easier to identify on this imagery due to distinctive red and dark tones for red pine and jack pine, respectively. Red oak was distinctive on the false color composite due to its brighter red color. The mixed hardwoods could also be easily identified on the false color imagery. No distinction between private and government property was done on the 1:100,000 imagery.

3.5.2.2 **Second Stage Sampling Selection** The selected primary units on each of the three strata were transferred to the 1:24,000 CIR aerial photographs, which constituted the second stage. The dimensions of the primary units when transferred to the CIR photographs were rounded to one decimal place. As a result, the primary units on the CIR photographs had a smaller area (2,518 acres) since they were drawn 13.3 cm on a side. The difference in area is small (0.5%) and can be considered negligible.

The primary units were then divided into sixteen secondary units, each with an area of 157.4 acres. Two of these secondary units were selected within each previously selected primary unit. The selection process for the secondary units, and the size variable used to calculate the selection probabilities, were the same as for the primary units.

At this stage, it was possible to consider the private properties that were located within the secondary units.

Whenever a forest type was located within the limits of private property, no area measurement was done.

3.5.2.3 Third Stage Sampling Selection Each selected secondary unit was divided into one hundred and twenty one tertiary units of 1.3 acres. Three tertiary units were selected within each previously selected secondary unit. The selection process was the same as described above, but the size variable used for the calculation of the selection probabilities was percent crown closure. Anderson (1979) and Lee _et al_. (1984) made use of this size variable for the calculation of their selection probabilities. The crown closure estimates were photointerpreted using a crown-density scale which was graduated in 10% intervals from 5% to 95%. The assumption, of course, is that higher percentage crown closure within a tertiary unit reflects higher timber volume.

This assumption is not necessarily true because a high value of crown closure is not always associated with a high timber volume per unit area. The observer may be looking at a dense population of thin trees, such as is found in overstocked young stands. Aerial volume tables that use percentage crown closure as an independent variable must have another independent variable, such as average total height of trees. Such a table is available for the northern mixed forest types (Avery _et al_. 1959). It was not used because of the impossibility of estimating total height from the airphotos used, since the ground could not be seen in the points where the tertiary units were located.

The ideal situation would be to take the CIR photographs showing the location of the tertiary units selected into the field. This was not possible because they were available for office work only. The tertiary units were therefore transferred from the CIR photographs to base maps of each township using a reflecting projector. The maps with the location of the tertiary units were taken to the field.

3.5.3 Multiphase Sampling Technique

3.5.3.1 Digital Image Processing As already stated, the multiphase sampling technique to be evaluated by this research required the use of digital imagery for the study area. In this section only an overview will be presented of the steps taken to digitally process the subscene of the study area. For more detailed treatments of digital image processing, the reader should consult specialized literature on the subject such as Lillesand and Kiefer (1987), Hudson (1986) and Jensen (1986).

The digital processing of the image was performed on a micro-computer from the Center for Remote Sensing. A Landsat digital analysis software package was used. This is a modular interactive software system that allows the user to display a color image composed of three bands within a 240 by 256 pixels (picture element) subscene (Hudson, 1986).

In order to digitally process the Landsat image, a subscene (240 x 256 pixels) of the study area had to be created. With the computer compatible tape (CCT) loaded into the tape drive unit of the ERDAS-400, the study area was located using the READBIL program. This program searches the CCT for a specific area of interest based on the coordinates of the center of the area. These coordinates are found by aligning a transparent grid over the Landsat image to show the row and column numbers for the area of interest and the center coordinates.

Once the subscene containing the study area was located, a file (with extention .LAN) was created using the program LOADBIL. These two programs are utilized with CCT's written in the band-interleaved-by-line (BIL) format. In this format, the data for all the bands are written line by line onto the tape i.e., line 1 band 1, line 1 band 2, line 1 band 3, line 1 band 4, etc. This format is useful if all bands are to be used in the analysis (Jensen, 1986).

Once the .LAN file was formed, it was necessary to calculate some fundamental univariate statistics for subsequent processing of the data file (Hudson, 1986; Jensen, 1986). These statistics were created by the BUILDH program which has two functions: a) to add or modify information contained on the header (e.g. number of rows, columns and bands; (X,Y) coordinates for the upper left corner; etc) or the trailer (e.g. histogram; mean; mode; standard deviation; minimum and maximum brightness values;

user assigned band number) and b) to compute statistics describing the image.

To display the image the READ program was used. This program allows the display of up to three selected bands of a .LAN file as an intensity-modulated image of red, green or blue, superimposed to create a false color composite image (Hudson, 1986). The image can be displayed, magnified or reduced.

The image displayed on the screen was not geometrically rectified to correct for skew (i.e. earth-rotation distortion). The process of geometric rectification is necessary for some applications that require accurate measurements of area, direction or distances. As this was not the case with the present study, it was decided not to do the rectification. Besides, in the process of geometric rectification of an image, the exact geometric relationship between the input pixel location (row, column) and the associated map coordinate of the same point is rarely obtained and the brightness values are also modified (Jensen, 1986).

The subscene obtained from the CCT encompassed an area larger than the study area. The CURBOX program was used to define the study area within the subscene extracted from the CCT. A rectangular box is drawn on the screen and the upper left and lower right corners of the study area are defined. Locating the corners of the study area on the displayed image was very difficult because of the resolution of the image. The CIR photographs were used to help identify them as precisely as possible.

After the (X,Y) coordinates of the two diagonally-opposite corners of the study area were defined, the SUBSET program was used to copy the study area to a new file. This program fills the part of the image surrounding the study area with zeros, thereby eliminating the possibility of unwanted data on the new file. The statistics for the study area were obtained after it was isolated within the subscene.

Once the new file was formed, the READ program was used again to display the image. An enhancement program called HISTOEQ, which performs a histogram equalization of the displayed image was used. This program enhances the image by maximizing the color contrast.

Before the classification was performed, the PRINCE program, which calculates the principal component of the data, was used. The number of principal components is equal to the number of bands on the original data. In the present case, there were four principal components since the original data had four bands.

To do the classification, the MAXCLASS program was used. This program performs a supervised classification of a .LAN data file by either the minimum distance-to-means or the maximum likelihood classifiers. The selection of supervised classification was done based on the study by Hudson (1986) which found that unsupervised classification, on a site with similar forest cover as the study area, gave low classification accuracies as compared to supervised classification. The same author also found that minimum distance-to-means was more accurate than the maximum likelihood algorithm for the same test area referred to above.

Although these results were obtained using a winter scene, as opposed to a fall scene as in the present study, the minimum distance-to-means classifier was used to classify the study area. Another point that contributed to the selection of this classifier is the fact that it takes less computer time to do the classification than the maximum likelihood algorithm.

The minimum distance-to-means classifier calculates the spectral distance between each pixel and the mean value for each training site. The pixel is assigned to the class for which the distance is smallest (Hudson, 1986).

A requirement of the MAXCLASS program is that a signature file be created. This file was created by using the program called FIELD. This program was used to build and save, in a signature file, the statistics of training sites within the study area.

By using a joystick, a polygon was drawn on the screen which encompassed an area of known cover type. Based on the histogram and on the summary statistics shown on the screen, a total of thirty three training sites were obtained. The training stage for supervised classification is critical since the success of the classification relies on it, and is more of an art than a science (Lillesand and Kiefer, 1987). The CIR photographs were used to help in the selection of the training sites. The training areas were selected independently on both townships that constitute the study area, but the classification was done on the whole area.

Seventeen training sites were initially obtained from Henderson Township (T.21N. R.11W.) on the following cover types: red pine; jack pine; red oak; mixed hardwood; agriculture and water. From Boon Township (T.22N. R.11W.), sixteen training sites were obtained for all the same types except red oak, since it does not occur in the area.

The program SIGNAM was used to create a list of signature names which were stored on a file of signatures. This file is utilized by MAXCLAS. The term "signature" refers to the spectral response measured by remote sensors over earth targets. The term "signature" as used in remote sensing literature, implies a pattern that is absolute and unique. The spectral patterns observed in remotely sensed data may be quantitative, but they are not absolute; they may be distinctive, but they are not unique (Lellesand and Kiefer, 1987).

Since there were several training sites for the conifer, mixed hardwood, red oak and agriculture strata, it was decided to synthesize a new class from several of the ones initially obtained for those cover types. The ADDSIG program was used to accomplish this. This program allows the combination of several signatures from a given file to be added to the end of the file. The new synthesized classes were as follows: two red pine classes were synthesized from five; one jack pine from four; two red oak from four; two mixed hardwoods from five; and two agriculture from five. The difference between the new synthesized classes and the old classes, is that the former have a broader range of brigtness values as a result of the combination of the classes. The classification was done by using the new synthesized classes.

The classified image was displayed on the screen by using the DISPLAY program. The difference between the DISPLAY and the READ program is that the former is used when a file with a .GIS extention is to be displayed, which is the case for the classified image. The later program is used when a file with a .LAN extention is involved.

In order to obtain a hard copy in the form of a number map of the study area, the PUPDATE program was used to eliminate the zeros that surrounded the area. After this, the NPRINT program was used to print the number map from the GIS file. The hard copy or number map shows how each pixel in the data file was classified and its respective coordinates. As a result, it was possible to delineate within the number map the clusters of pixels that belonged to the cover types of interest.

The number map with the delineated clusters of pixels was then used in the selection process of the sample. To locate the selected pixels (sampling units) on the image, the CURSES program was used. This program displays a cursor on the screen and allows the user to move it around the scene. The screen and data file coordinates of the selected sampling units were shown on the screen, allowing them to be located within the study area.

3.5.3.2 Phase one Sample Selection The whole population was divided into four strata: conifer, mixed hardwood, agriculture and water. Each stratum had a known number of sampling units (pixels), which were obtained from the classified image. The number of units for each stratum is presented in Table 1.

Table 1

Total Number of Sampling Units on Each Stratum

Stratum	Sampling Units (# of pixels)
Conifer	8,538
Mixed Hardwood	35,929
Agriculture	12,775
Water	213

Only the conifer and mixed hardwood strata were of interest in the present study since the objective was to estimate wood volume. As a result, the non-forest strata agriculture and water were not considered. The towns located within the area were included on the non-forest stratum, and classified as such, since they were of no interest for the objectives of the study. For the two strata of interest, sampling fractions were arbitrarily chosen so as to allocate a reasonably large number of sampling units to each stratum (Johnston, 1987). This was done because no previous information on the use of the method was available for the study area.

Sampling fractions of $u_1 = 0.0144$ and $u_2 = 0.00315$ were chosen for the conifer and the mixed hardwood strata, respectively. By entering the values of the u_i and N_i ($i = 1, 2$) from Table 1 in equation 10, the following number of phase one sampling units were obtained: conifer stratum (n_1)= 123; mixed hardwood stratum (n_2) = 113.

At this point, each stratum was further stratified into two (I_i) phase one strata. The conifer stratum was divided into red pine and jack pine phase one strata. The mixed hardwood stratum was divided into hardwood and red oak phase one strata.

The samples (n_1 and n_2) within each stratum were randomly selected and classified into one of the two phase one stratum. The selection process was done with the help of the number map. Previous to this selection process, the cover types of interest were identified and delineated on the number map. The CIR photographs were used to help identify and locate the cover types. The coordinates (row and column numbers) of these cover-type clusters were obtained from the number map in order to define their positions within the study area.

The selection process of the sampling units used a table of random numbers, from which a pair of numbers was selected. The first number of the selected pair was assigned to a row on the number map and the second number to a column. Since the study area was composed of 386 rows and 262 columns, the selection of the pair of numbers, which defined the (X,Y) coordinates a sampling unit was done only within those limits. Selected sampling units that were located within the boundaries of private property were discarded. The CIR photographs were used to identify the parcels of private property and to determine whether a selected sampling unit was within their limits.

The selection process resulted in the following distribution of sampling units per phase one stratum: from the n_1 = 123 conifer units, n_{11} = 23 were jack pine and n_{12} = 100 were red pine; from the n_2 = 113 mixed hardwood units, n_{21} = 10 were red oak and n_{22} = 103 were hardwood. These results are in agreement with equation 11.

3.5.3.3 Phase two Sample Selection At phase two, a subsample of size m_{ij} given by equation 13 is selected. The sampling fractions v_{ij} for i=j = 1,2 were again arbitrarily chosen for each stratum in order to obtain a sample of five units per stratum, except for the red oak in which six samples were measured due to its small size.

Table 2 presents the sampling fractions v_{ij} for each of the stratum considered.

Table 2

Values of the Sampling Fraction

Stratum	v_{ij}
Jack pine	0.22
Red pine	0.05
Hardwoods	0.05
Red Oak	0.60

The second phase sampling units were selected using a random-number table by selecting five numbers in the intervals between 1 and each n_{ij}.

3.5.3.4 Location and Transfer of the Samples from the Imagery to the Base Map

The location of the selected samples on the digital representation of the imagery was done on the computer by using the CURSES program, as already stated. Since the number map shows the coordinates of each sampling unit, their location on the imagery was relatively easy.

Once a selected sampling unit was located in the imagery, the CIR photographs were consulted in order to better identify its location in the study area. This was done by calculating the distance of the sampling unit to two distinct features on the imagery, such as road intersections. The distance was determined by counting the number of pixels to the reference points. As the imagery was not geometrically corrected, the pixels were considered as representing an area of 56 x 56 meters square (Lusch, 1988). After finding the "coordinates" of the sampling unit relative to some feature in the image, these values were transferred to the CIR photographs to identify the location of the cover type and to the base map by converting the calculated distances to the appropriate scale.

3.5.4 Equations used for Volume Estimation In order to estimate the volume of individual trees considered as "in" during the field measurements, volume equations developed for the Huron-Manistee National Forest by the US Forest Service were used (Hesse, 1987). The general form of the equation is:

$$v_i = A(1 - e^{(C * DBH)})^B \qquad (20)$$

where: v_i = volume of the i^{th} tree in cubic feet

A, B and C are coefficients

DBH = diameter at breast height in inches.

This equation gives net volume without bark to 4 inch minimum top diameter.

Table 3 shows the regression coefficients for the above equation, for the species found during the field work. Appendix C presents the scientific name of the trees.

Table 3

Coefficients for the Volume Equation by Species

Species	A	B	C
Red pine	119.83517	4.10017	-.08516
Jack pine	100.91007	4.15180	-.0878
White oak	120.63656	3.99791	-.07755
Red oak	185.60448	3.78645	-.06456
Aspen	81.04818	4.90735	-.1164
Sugar maple	49.16485	6.25352	-.1631
Red maple	105.78777	4.19832	-.09
Cherry	160.14361	3.71334	-.0668
Beech	161.83706	3.82825	-.0675

For the species not listed on Table 2, the following general volume equation presented by Beer *et al.* (1966) was used:

$$v_i = (D^2(D+190)/100{,}000)((H(168-H)/6400)+(0.32/H))(79) \tag{21}$$

This equation gives the volume of the stem in cubic feet, without bark, from a 1 foot stump to a point on the bole where merchantability is limited by branches, deformity or minimum diameter (Beers, 1964). In the present study, the minimum diameter used was 4 inches.

This general volume equation was originally developed for estimating volume in standard cords. In order to achieve greater flexibility in its usage, merchantable height was substituted for the number of 8 foot bolts and the conversion factor 79 (to convert from standard cord to cubic feet without bark) was introduced (Beers, 1964).

For the calculation of the volume per unit area from each sample point within a sampling unit, the following equation was used (De Vries, 1986):

$$V_j = k \sum_{i=1}^{N} v_i / g_i \tag{22}$$

where: V_j = volume per acre for the j^{th} point sampling

k = basal area factor (BAF = 10)

v_i = volume of the ith "in" tree given by one of the above equations, in cubic feet

g_i = basal area of the i^{th} "in" tree, in sq ft

N = total number of "in" trees with DBH > 4in.

The final estimate of the volume per unit area for a sampling unit was obtained by averaging the volume per unit area for the two point sampling.

3.5.5 Area Estimation The measurements for area estimation of the cover types within the study area were done using an electronic planimeter. The CIR photographs with the overlays delineating the forest types attached to them were used. The specific planimeter used allowed for the addition or subtraction of the readings done in different parts of the CIR photograph without the necessity of readjustments. This characteristic of the equipment helped in this phase of the work by accelerating the process without loosing the precision of the readings.

Some sections of the boundary of the study area were not very obvious on the CIR photographs. As a result, the total area of the study area was estimated by using the base maps. As these maps also show the limits of the private property, they were used to estimate the areas of those properties.

3.5.6 Time Measurements and Costs The number of hours required for each activity was recorded in order to estimate the total cost for each sampling technique evaluated. The individual activities were: photointerpretation, computer classification during digital processing, sampling selection, area measurements, and field work.

Since the time was measured in terms of hours spent on each activity, an estimation of the price per hour that a consulting forester would charge to do the same type of work would be necessary. To estimate this price, a telephone survey was done on a random sampling of registered consulting foresters in the State of Michigan. It was assumed that the price per hour included mileage. The average price obtained from this survey was used to estimate the costs involved for each sampling procedure.

The cost for purchasing imagery (CIR photographs and satellite images) was not included since these images were available for use from the Michigan DNR and the Center for Remote Sensing. The price per hour for the use of the microcomputer was, however, included. It is based on the rate established by the Center for Remote Sensing for non-University clientele. For the purpose of this study, it was assumed that the computer would be used only during normal office hours, and that no operator would be required.

CHAPTER 4

RESULTS

This chapter presents the results obtained from the application of the sampling techniques selected to the study area. No discussion of the results are made in the present chapter as these follow in Chapter 5.

First the results obtained from the multistage sampling technique are presented. These include the estimated total volume and its standard deviation, the coefficient of variation and confidence intervals for the estimated population total and for the estimated total for each stratum separately. The results obtained from the sensitivity analysis are presented in several tables and also in graphical form. The relationship between the size variable and the estimated volume for the second and third stages are presented in graphical form for each stratum. The estimated basal area and acreage for each stratum are also shown.

For the multiphase technique, the results include the numbers obtained from the digital processing of the imagery, the estimated total volume and its standard deviation for the population and for each stratum. The estimated basal area for each stratum are also presented. Time spent for each activity and the respectives costs are shown for both sampling techniques.

4.1 Multistage Sampling Technique During the field work on the conifer and red oak strata, it was noticed that the stands had been thinned. Information obtained from the regional office of the U.S. Forest Service, located in Cadillac, Michigan, showed that the red oak stand had been thinned during 1986 - 1987. The conifer stands, including red and jack pines, were thinned in 1986 (Norton, 1987).

This thinning, of course, did not show on the CIR photographs which were taken in 1977. As a result, the probabilities for the selection of the tertiary sampling units had to be recalculated since they were based on the percent crown closure.

In order to recalculate the probabilities, newly-acquired 1:15,840 black and white infra-red photographs taken in 1987 were used. The overlays originally used for the 1:24,000 CIR photographs were enlarged to the scale of the new photos and the same crown density scale was used to estimate percent crown closure. The estimated total volume and its variance for the red oak and conifer strata were recalculated based on the new probabilities.

The estimated total volume, its standard deviation, the coefficient of variation and the confidence intervals for the three strata and the population, are presented in Table 4. For the multistage technique, the population estimate was obtained by the summation of the estimated total volume per

Table 4

Estimated Total Volume, Standard Deviation, Coefficient of Variation and Confidence Intervals for Each Stratum and the Population

MULTISTAGE

Stratum	Total Volume (cuft)	Standard Deviation	C.V.(%)
Conifer	131,112,680	4.13866E+7	31.6
Hardwood	234,377,358	6.92547E+7	29.6
Red Oak	14,082,693	1.29223E+7	91.8
Population	379,572,731	8.17072E+7	21.5

Confidence Intervals (cuft)

Conifer	[48,339,405 , 213,885,955]
Hardwood	[95,867,927 , 372,886,789]
Red Oak	[0 , 39,927,371]
Population	[219,158,447 , 542,987,014]

MULTIPHASE

Stratum	Total Volume (cuft)	Standard Deviation	C.V.(%)
Conifer	115,621,019	1.36771E+7	11.8
Hardwood	93,652,660	4.02768E+7	43.0
Red Oak	104,316,573	2.54317E+7	24.4
Population	76,336,702	1.04770E+7	13.7

Confidence Intervals (cuft)

Conifer	[88,266,838 , 142,975,200]
Hardwood	[13,099,005 , 174,206,314]
Red Oak	[53,453,232 , 155,179,913]
Population	[55,382,703 , 97,290,699]

stratum (Langley, 1975). The variance of the total was calculated by summing the variance for each stratum. This was done based on the property that the variance of the sum of K independent variables is equal to the sum of the variances of the K variables (De Vries, 1986).

Table 5 presents the total and merchantable basal area for the three strata. The total basal area for the conifer stratum refers to all trees considered as "in" on both points, including hardwood species scattered among the pine trees. Merchantable basal area refers to only conifer trees (red pine and jack pine). For the hardwood and red oak strata, the total basal area values refer to all trees considered as "in" on both points, independent of their diameter at breast hight (DBH). Merchantable basal area refers to trees with DBH greater than 4.0 inches. All the values of basal area shown represent the average of twelve sampling units, with two points per unit. The number within the parentheses express the range of the basal area found on the plots measured on each stratum.

Approximate 95% confidence limits for the total volume for each stratum can be determined by $\hat{y} \pm t_{(0.95)} \sqrt{var(y)}$. This requires the following assumptions (De Vries, 1986): ignoring the number of degrees of freedom on the variance of the total; assuming that the estimated total is normally distributed about its parameter with variance given by squaring the standard deviation and setting $t \approx 2$. These

Table 5

Total and Merchantable Basal Area
for Each Stratum

Stratum	Total (sqft/acre)	Merchantable
Conifer	120 (85 to 180)	117 (70 to 180)
Hardwood	122 (75 to 150)	112 (75 to 145)
Red Oak	116 (85 to 160)	111 (85 to 160)

confidence limits are also shown in Table 4 for each of the stratum and for the population.

Table 6 shows the estimated area in each stratum for the government and private properties and the total for the study area. The values for the areas of each stratum do not sum to the government total area since those values represent the net areas. It means that the areas not covered with a forest type were subtracted during the process of measurement on the CIR photos. The government and private land figures add to the total since these readings were obtained from the base maps.

The estimated volume per acre and its standard deviation for each stratum, and for the population, are shown in Table 7. To calculate the estimated volume and the variance per acre, total volume and its variance were divided, respectively, by the total area and the total area squared (shown in Table 6). Appendix E presents the volume per acre for each plot measured on the three strata.

Table 6

Estimated Areas for Each Stratum, Government and Private Land and Total

Stratum	Area (acres)
Conifer	7,783
Hardwood	16,349
Red oak	1,529
Government	32,203
Private	15,649
Total	47,852

Table 7

Estimated Volume per Acre and Standard Deviation

	Conifer	Hardwood	Red Oak	Population
Volume (cuft/acre)	2,740	4,898	294	7,932
Std. Dev.	864.888	1,716.333	270.048	1,707.495

Table 8 presents the time, in hours, spent on the activities related to the multistage sampling technique and the respective costs. The average price per hour obtained from the survey of selected consulting foresters in Michigan was $29.00.

Within the parentheses, the number on the left represents the total time required to measure the twelve plots selected on each stratum. The right-hand number within the parentheses refers to the total time spent moving between plots within each stratum and the time required to reach the first plot measured on that day. The times shown for the first, second and third stages for each stratum refer to the measurement and selection of the sample units in those stages.

4.1.1 Proportion Between Measure of Size and Predicted Volume

The proportionality between the size variable and the variable of interest in sampling with probability proportional to size is an important characteristic. The best measure of size for calculating the selection probabilities would be the one that is proportional to the variable of interest (Cochran, 1977; De Vries, 1986):

Table 8

Time Spent on the Activities for the Multistage Technique

Activity	Time (hours)	Cost ($)
Photointerpretation	18.3	530.70
Sampling Measurement		
Conifer		
First stage	15	435.00
Second stage	10.6	307.40
Third stage	13	377.00
Field	16.7 (9.5+7.2)	481.40
Hardwood		
First stage	10.5	304.50
Second stage	1.5	43.50
Third stage	22.3	646.70
Field	27.4 (14.3+13.1)	794.60
Red oak		
First stage	5.8	168.20
Second stage	12	348.00
Third stage	17.8	516.20
Field	23.7 (9.1+14.6)	687.30
Area Measurements	41.3	1,197.70
Total	235.9	$6,841.10

$$p_i = y_i / Y$$

where y_i is the size of the variable of interest and Y is the total for the variable of interest.

This implies that the population total, the estimate being sought, is known in advance. In practice, it is common to work with probabilities that are proportional to an easily assessable measure of size (x_i), one that should also be well correlated with the variable of interest. The probabilities calculated this way are approximations of the optimum probabilities that would minimize the variance of the estimator. However, the higher the correlation between the chosen measure of size and the variable of interest, the more the calculated probabilities will approximate the optimum probabilities. The end result will be a smaller variance of the total (Langley, 1975; Anderson, 1979; De Vries, 1986).

For each of the three strata, Figures 2 and 3 present the relationship between the size variable and predicted volume for the second and third stages respectively. As already stated, the size variables used for the second and third stages were estimated total area occupied by the corresponding forest type within the sampling unit, and percent crown closure, respectively. Also shown on these graphs are the simple correlation coefficients between the two variables. No relationships were developed for the first stage, since only two primary units were selected on each stratum.

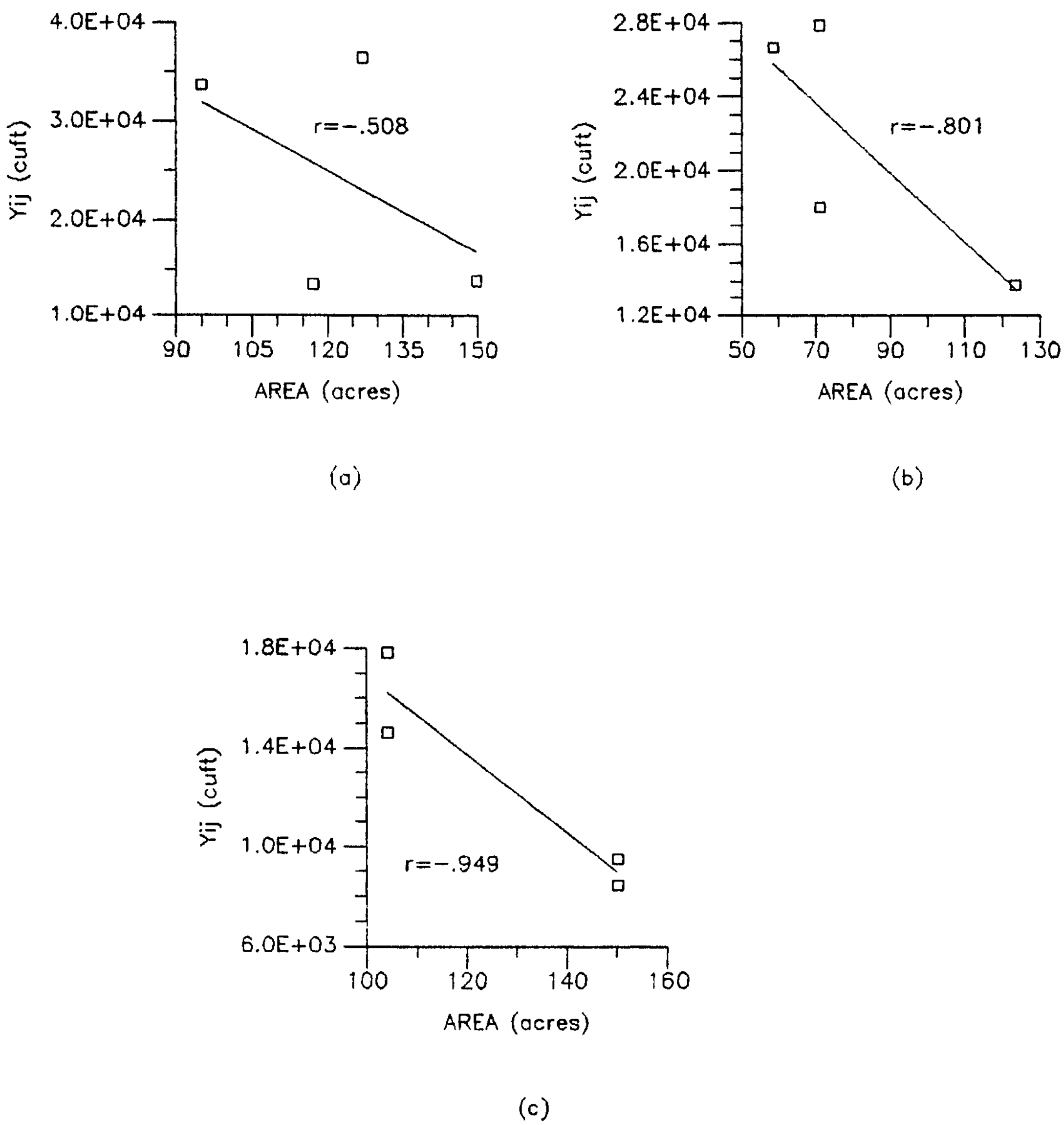


Figure 2 Prediction Variable versus Predicted Volume for the Second Stage for a) Hardwood b) Conifer and c) Red Oak

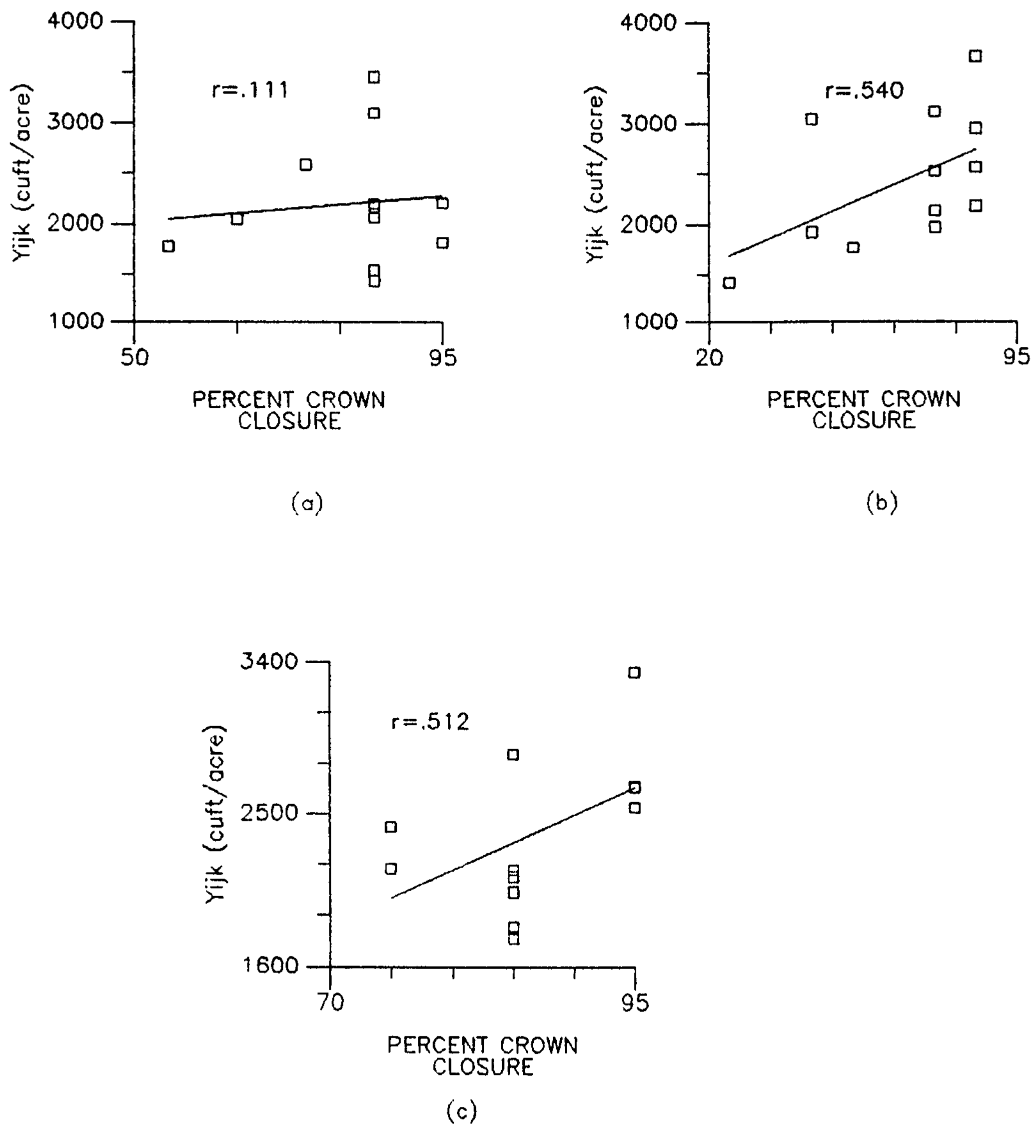


Figure 3 Prediction Variable versus Predicted Volume for the Third Stage for a) Hardwood b) Conifer and c) Red Oak

4.1.2 Sensitivity Analysis

Table 9 presents the estimated total volume and its variance for each of the five trials of the conifer stratum with changes in the selection probabilities. The selection probabilities (the actual values are presented in Appendix B) are: P_i for the first stage; P_{ij} for the second stage and P_{ijk} for the third stage.

Table 10 shows a summary of the effects of changing one specific selection probability, keeping the others as originally calculated. The percentage shown on the upper portion of the letters D (for decrease) or I (for increase), refers to the change that occurred in the estimated total volume. The percentage shown on the lower portion, refers to the change that occurred in the estimated variance of the total.

Table 11 presents the estimated total volume and its variance for each of the five trials for the hardwood stratum with changes in the selection probabilities. These probabilities are presented in Appendix B.

Table 12 shows a summary of the effects of changing a specific selection probability while keeping the other two as originally calculated. The percentage shown on the upper portion of the letters D (for decrease) or I (for increase), refers to the change that occurred in the estimated total volume. The percentage shown on the lower portion, refers to the change that occurred in the estimated variance of the total.

Table 9

Changes in the Estimated Total Volume and its Variance with Changes in the Selection Probabilities for Conifers

Probability	Trial#	Volume (cuft)	Variance
	1	114,368,973.46	1.50015E+15
	2	142,900,588.66	1.86155E+15
P_i	3	120,958,301.38	1.58407E+15
(first stage)	4	141,564,788.60	1.84474E+15
	5	119,531,602.72	1.56592E+15
	1	133,326,981.90	1.77107E+15
	2	132,174,896.31	1.74086E+15
P_{ij}	3	135,191,295.14	1.82094E+15
(second stage)	4	127,151,392.80	1.61084E+15
	5	118,945,960.84	1.40974E+15
	1	138,831,787.85	1.92385E+15
	2	135,303,420.66	1.82587E+15
P_{ijk}	3	133,761,233.23	1.78387E+15
(third stage)	4	145,720,643.25	2.12247E+15
	5	128,288,304.01	1.63872E+15

Table 10

Effects of Changing the Selection Probabilities on the Estimated Total Volume and its Variance for Conifer

Probability	Trial# 1	2	3	4	5
P_i	14.6% D 14.2%	9.0% I 8.7%	8.4% D 8.1%	8.0% I 7.7%	9.7% D 9.4%
P_{ij}	1.7% I 3.4%	0.8% I 1.6%	3.1% I 6.3%	3.1% D 6.3%	10.2% D 21.5%
P_{ijk}	5.9% I 12.3%	3.2% I 6.6%	2.0% I 4.1%	11.1% I 23.9%	2.2% D 4.5%

D = Decrease I = Increase

x% (percent change in estimated total volume)
D
y% (percent change in estimated variance of the total)

Table 11

Changes in the Estimated Total Volume and its Variance with Changes in the Selection Probabilities for Hardwood

Probability	Trial#	Volume (cuft)	Variance
	1	222,257,862.32	4.55412E+15
	2	251,759,795.15	5.32574E+15
P_i	3	222,399,384.29	4.55690E+15
	4	213,690,995.25	4.38260E+15
(first stage)	5	254,642,391.44	5.19950E+15
	1	218,435,668.84	4.16542E+15
	2	218,623,101.65	4.17066E+15
P_{ij}	3	240,506,803.61	5.05065E+15
	4	236,512,959.40	4.88136E+15
(second stage)	5	216,074,784.30	4.07536E+15
	1	245,628,425.85	5.27383E+15
	2	238,970,339.44	4.98846E+15
P_{ijk}	3	245,690,373.36	5.27652E+15
	4	236,115,681.76	4.86853E+15
(third stage)	5	264,367,392.86	6.11962E+15

Table 12

Effects of Changing the Selection Probabilities on the Estimated Total Volume and its Variance for Hardwood

Probability	Trial# 1	2	3	4	5
P_i	D 5.5% / 5.3%	I 7.4% / 11.0%	D 5.4% / 5.3%	D 9.7% / 9.4%	I 8.6% / 8.4%
P_{ij}	D 7.3% / 15.0%	D 7.2% / 15.0%	I 2.3% / 5.3%	I 0.9% / 1.8%	D 8.5% / 17.7%
P_{ijk}	I 4.8% / 9.9%	I 1.9% / 4.0%	I 4.8% / 10.0%	I 0.7% / 1.5%	I 12.8% / 27.6%

D = Decrease I = Increase

D x% (percent changes in estimated total volume) / y% (percent changes in estimated variance of the total)

Table 13 presents the estimated total volume and its variance for each of the five trials for the second and third stages of the red oak stratum achieved with varying selection probabilities. The selection probabilities are shown in Appendix B. The reduced number of first stage units contained in this stratum resulted in only three trials, that did not repeat the same value for the selection probability as calculated for other trials (see section 3.1.1).

Table 14 shows a summary of the effects of changing a specific selection probability and keeping the other two as originally calculated. The percentage shown on the upper portion of the letters D (for decrease) or I (for increase), refers to the change that occurred in the estimated total volume. The percentage shown on the lower portion, refers to the change that occurred in the estimated variance of the total. The effects for trials #2 and #3 for the probability P_i are not shown since values of these probabilities were equal to the probabilities P_i originally calculated. For trial #5, the probabilities were equal to those of trial #4.

The results of changing the selection probabilities on the estimated total volume and its variance for all of the combinations, for each stratum, can be seen from the plots shown on Figures 4, 5 and 6. The letters A, B, C, D, and E on Figures 4 and 5 and A, B and C on Figure 6, refer to the groups of combinations.

Table 13

Changes in the Estimated Total Volume and its Variance associated with Changes in the Selection Probabilities for Red oak

Probability	Trial#	Volume (cuft)	Variance
	1	13,683,197.60	1.62261E+14
P_i	2	14,082,692.81	1.66987E+14
(first stage)	4	14,571,130.86	1.72763E+14
	1	13,234,186.48	1.47472E+14
	2	14,487,685.39	1.76724E+14
P_{ij}	3	13,385,862.14	1.50872E+14
	4	13,792,302.19	1.60168E+14
(second stage)	5	15,652,432.94	2.06284E+14
	1	14,510,836.58	1.77307E+14
	2	13,936,557.96	1.63536E+14
P_{ijk}	3	14,456,470.97	1.75979E+14
(third stage)	4	13,889,661.25	1.62435E+14
	5	14,679,645.13	1.81461E+14

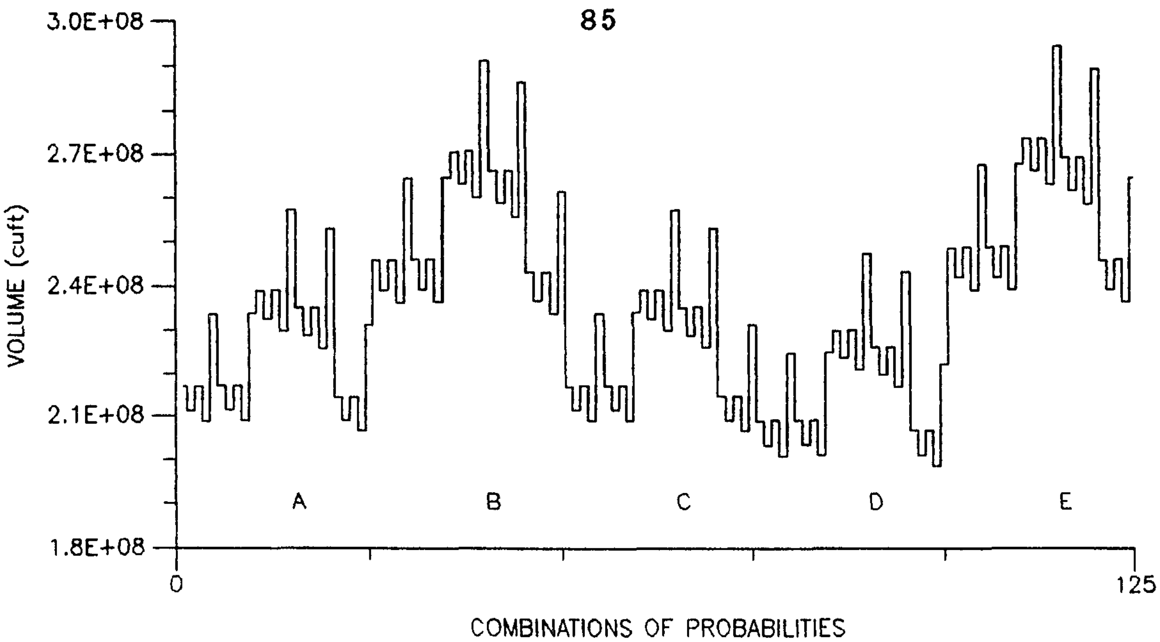
Table 14

Effects of Changing the Selection Probabilities on the Estimated Total Volume and its Variance for Red Oak

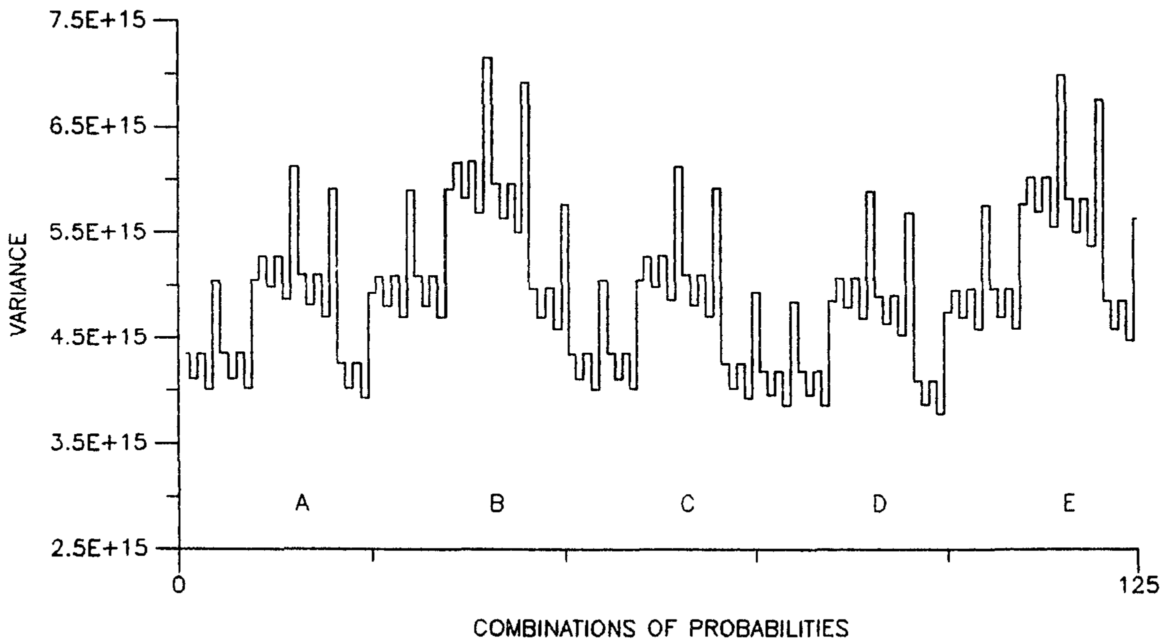
	Trial #				
Probability	1	2	3	4	5
P_i	2.9% D 2.9%	NC	NC	3.5% I 3.5%	same as trial 4
P_{ij}	6.4% D 13.2%	2.9% I 5.8%	5.2% D 10.7%	2.1% D 4.2%	11.1% I 23.5%
P_{ijk}	3.0% I 6.2%	1.0% D 2.1%	2.6% I 5.4%	1.4% D 2.8%	4.2% I 8.7%

NC = no change D = Decrease I = Increase

x% (percent change in estimated total volume)
D
y% (percent change in estimated variance of the total)

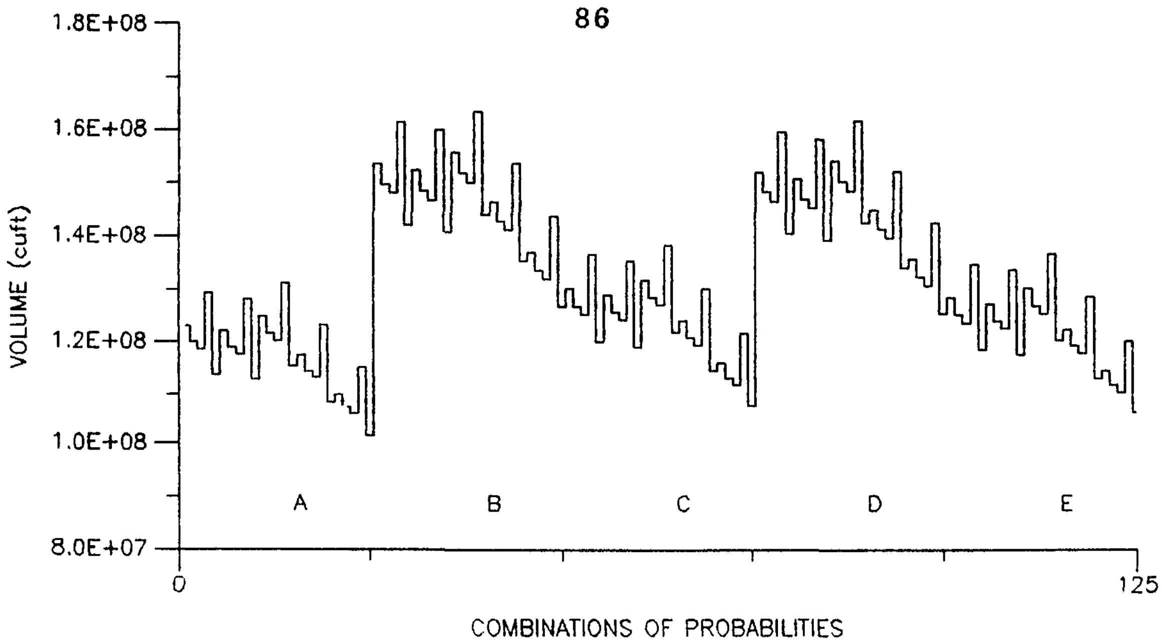


(a)

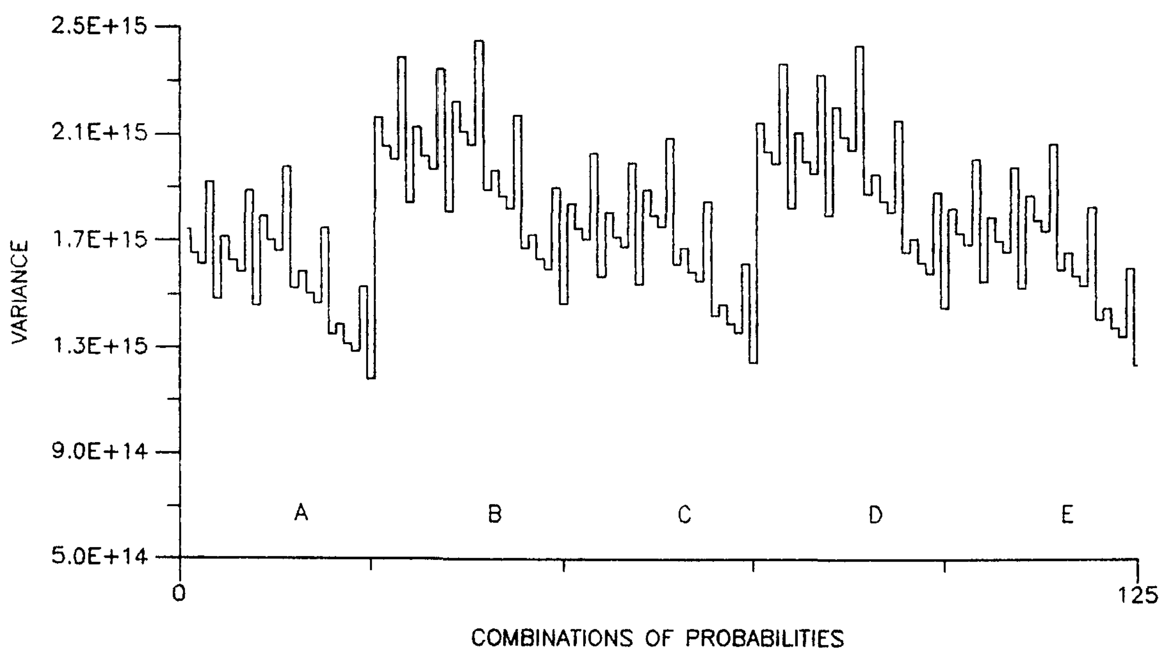


(b)

Figure 4 Effect of Changing the Selection Probabilities on a) Volume and b) Variance for Hardwood



(a)



(b)

Figure 5 Effect of Changing the Selection Probabilities on a) Volume and b) Variance for Conifer

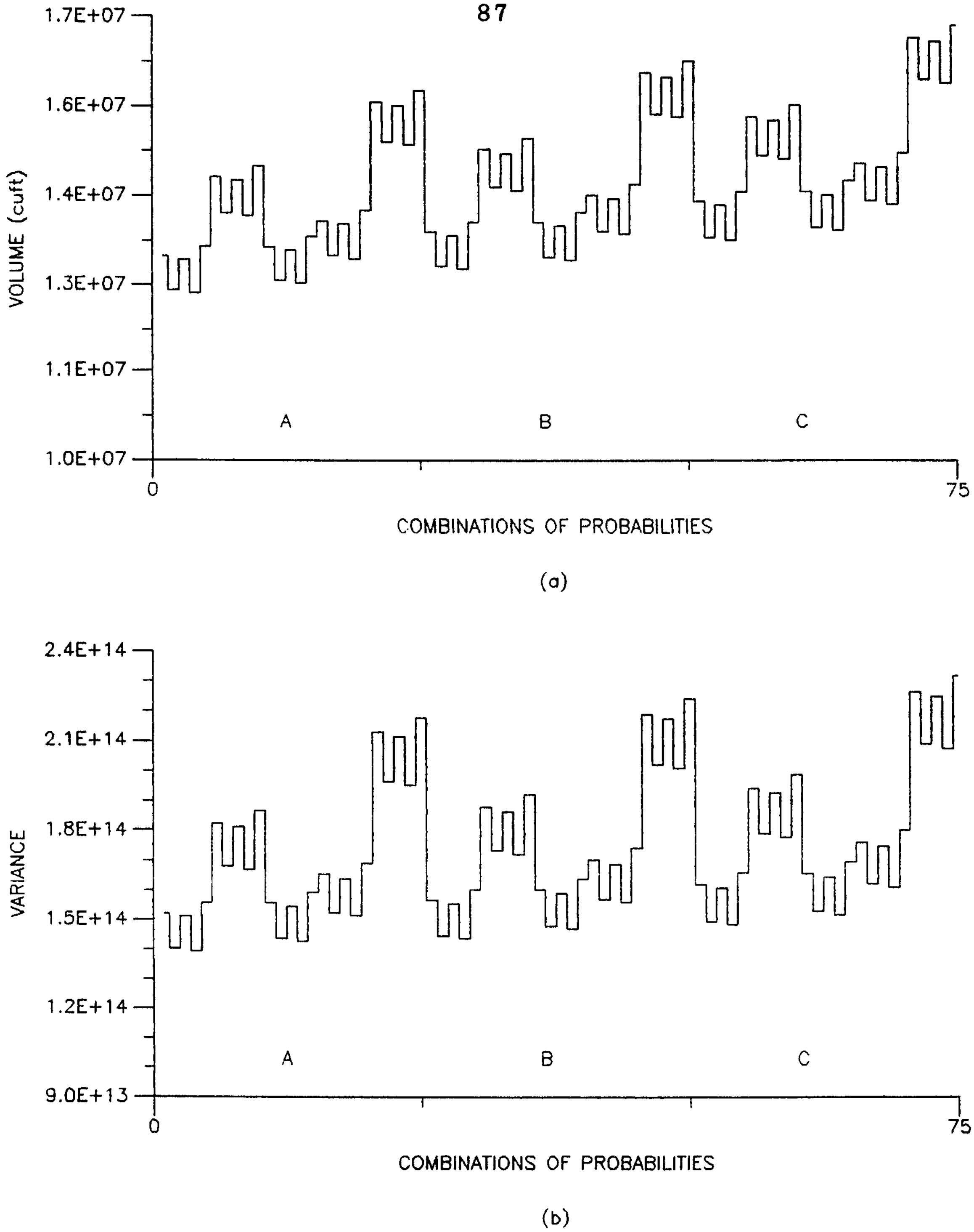


Figure 6 Effect of Changing the Selection Probabilities on a) Volume and b) Variance for Red Oak

Table 15 presents the minimum and maximum values for the total volume and its variance for each stratum, obtained as a result of all the combinations of the selection probabilities.

4.2 Multiphase Sampling Technique

4.2.1 Digital Image Processing The univariate statistics for the digital representation of the study area, composed of 386 rows and 262 columns, are presented in Table 16.

Table 17 presents the results of the classification of the digital representation of the study area. The numbers shown are from the histogram listing for the GIS file generated after the classification was completed. The percentages on the third column were calculated based on non-zero points, which totaled 57,455.

4.2.2 Estimated Volume, Variance and Time Measurements

Table 18 presents the estimated volume per unit area, the standard deviation and the coefficient of variation for the i^{th} population strata (conifer and hardwood).

The estimated volume per unit area, the standard deviation and the coefficient of variation for the four phase one strata and for the population are shown on Table 19. Appendix E presents the volume per plot for each of the four strata.

The total volume and its standard deviation, the coefficient of variation and confidence intervals for the

Table 15

Minimum and Maximum Values for the Total Volume and the Variance

	Volume	
	Min.	Max
C	101,520,931.79	163,762,490.93
H	198,464,956.22	294,736,993.87
R	12,682,506.40	16,881,819.60

	Variance	
	Min.	Max.
C	1.18135E+15	2.453000E+15
H	3.78008E+15	7.154320E+15
R	1.39393E+14	2.319200E+14

C = Conifer H = Hardwood R = Red Oak

Table 16

Univariate Statistics for the Study Area

Bands	4	5	6	7
Mean	18.654	17.377	65.036	70.918
Std. Dev.	2.913	5.387	15.425	19.891
Median	18	15	66	71
Mode	18	14	84	92
Minimum	10	7	6	0
Maximum	47	70	101	114

Table 17

Results of the Supervised Classification of the Study Area

Classes	# of Points	%
Zero Points	43,824	0
Red pine	7,229	12.58
Jack pine	1,309	2.28
Hardwood	28,230	49.14
Red oak	7,699	13.40
Others*	12,988	22.60

*Includes pixels classified as non-forest areas

Table 18

Estimated Volume per Acre, Standard Deviation and Coefficient of Variation for the i^{th} Population Strata

Stratum	Volume (cuft/acre)	Standard Deviation	C.V.(%)
Conifer	2,416	285.820	11.8
Hardwood	1,977	343.471	17.4

Table 19

Estimated Volume per Acre, Standard Deviation and Coefficient of Variation for the Four Phase one Strata and the Population

Strata	Volume (cuft/acre)	Standard Deviation	C.V.(%)
Red pine	2,625	777.164	29.6
Jack pine	1,507	278.006	18.4
Hardwood	1,957	841.695	43.0
Red oak	2,180	531.465	24.4
Population	1,595	218.946	13.7

conifer, hardwood and red oak strata and for the population are presented on Table 4.

The values shown in Table 18 for the hardwood stratum refer to the second stratum into which the population was initially divided. The first stratum was conifers. The hardwood stratum was further divided into two phase one strata: other hardwoods and red oak. The numbers shown in Table 4 refer to these phase one strata.

In order to obtain the total volume and the variance of the total for each stratum and for the population, the volumes per unit area and their variances were multiplied, respectively, by the total area and the total area squared. The total area was used so that the results could be comparable with the multistage technique.

The approximate 95% confidence intervals for the total volume of the strata and the population shown in Table 4 were calculated by using the equation presented in section 4.1 and taking into consideration the same assumptions.

Table 20 presents the averages for the total and merchantable basal areas for the four phase one strata. The averages are based on a total of 10 sampling points per stratum (2 points per plot). The numbers in parentheses refer to the ranges of the total and the merchantable basal areas. For the red and jack pine strata, total basal area refers to all trees considered as "in", independent of their DBH or species. Merchantable basal area refers to only conifer trees with DBH above 4 inches. As in the case of the

Table 20

Total and Merchantable Basal Area
for the Phase one Strata

Strata	Basal Area (ft^2/acre) Total	Merchantable
Red pine	134 (75-185)	132 (75-175)
Jack pine	110 (90-130)	97 (80-130)
Hardwood	101 (70-155)	94 (65-150)
Red Oak	108 (65-140)	92 (65-120)

plots measured on the multistage sampling technique, few hardwood species were found among the conifers.

For the hardwood stratum, total basal area refers to all "in" trees, independent of their DBH. Merchantable basal area refers to all "in" trees with DBH above 4 inches.

For the red oak stratum, the total basal area refers to all "in" trees regardless of their DBH. For the merchantable basal area, only red oak trees with DBH above 4 inches were considered.

Table 21 presents the time, in hours, spent on the different activities and their respective costs for the multiphase sampling technique.

Table 21

Time Spent on the Activities for the Multiphase Technique

Activity	Time(hours)		Cost($)
Computer Class.	67.30		336.50
	Phase one Strata		
Conifer Stratum			
P.I.*	18.30		530.70
Sampling selec.	36.80		1,067.20
Sampling trans.	2.00		58.00
Hardwood Stratum			
Sampling selec.	27.80		806.20
Sampling trans.	2.00		58.00
	Phase two Strata		
Conifer Stratum			
Red pine			
Field	8.60	(3.8+4.8)	249.40
Area measur.	8.00		232.00
Jack pine			
Field	7.50	(4.0+3.5)	217.50
Area measur.	8.00		232.00
Hardwood Stratum			
Hardwood			
Field	15.40	(5.0+10.4)	446.60
Area measur.	16.00		464.00
Red Oak			
Field	7.30	(3.9+3.4)	211.70
Area measur.	9.30		269.70
Total	234.30		$5,179.50

*P.I. = Photointerpretation

Appendix A presents equations for the calculations of the sampling fractions u and v, considering proportional allocation, and for the calculation of sample size (n). As for the multistage method, the average price per hour considered was of \$29.00. The price per hour for the use of the computer in classifying the scene was of \$5.00 (Lusch, 1988).

On Table 21, the number to the left within the parentheses on the lines for field work refers to the total time spent measuring the plots. The number to the right in parentheses refers to the total time required to move between plots. The activity called sampling selection consisted of: delineation of forest types on the number map; establishment of coordinates for the conifer and hardwood stands; selection of the phase one samples (n_i); location of the selected samples on the LANDSAT imagery; final sampling selection (m_{ij}'s). Sampling transfer refers to the cartographic transfer of the selected field plots from the CIR airphotos to the base maps.

CHAPTER 5

DISCUSSION

The discussion of the results presented herein is separated by stratum. The first section deals with the red oak stratum, the second section covers the conifer stratum and finally the third section deals with the hardwood stratum.

Section 5.4 discusses the implications of the sensitivity analysis. Section 5.5 discusses the results obtained for the population estimates, from both sampling techniques evaluated.

Digital image processing was performed just for the multiphase sampling technique and a discussion of the aspects related to the classification of the imagery into the different strata is presented in section 5.6. Sections 5.7 and 5.8 discuss repectively, the aspects related to time measurement and the costs involved in each technique and to the number of samples to be measured.

5.1 Red Oak Stratum The results presented on Table 4 for the multistage sampling technique show a low estimated total volume for the red oak stratum as compared to the other strata. Being the smallest stratum that occurs in the area, as seen from Table 6, it contained only two primary sampling units.

The particular multistage sampling technique evaluated has the tendency to concentrate the sampling units at each stage on areas of higher value or more volume, since they would have a higher probability of selection (Langley, 1971). For the red oak stratum, one of the primary units contained most of the stratum area and had a high probability of being selected. In fact, this unit was selected twice, as the technique implies sampling with replacement. This means that each time the primary sampling unit was selected, a new independent selection of secondary units had to be done (Langley, 1975). As a result, the four secondary units had to be selected within the same primary unit.

Due to the characteristic of the method to concentrate sampling units on areas of higher volume, the two secondary units selected for the first draw of the primary unit were the same as for the second draw of the primary unit. These two secondary units had highest probability of selection (see Appendix B).

The selection probabilities are used to calculate the expansion factors to estimate the total volume. The expansion factors for the first and second stages are given respectively by the sum of the ratios $1/P_i*n_i$ and $1/P_{ij}*t_{ij}$ as shown in equation 5. Due to the high probabilities of selection of the primary and secondary units, the expansion factors for these stages had a relatively low value. When these factors were multiplied by the estimate of the population total given by the sum of Y_{ijk}/P_{ijk}, the result was a low estimated total volume.

Under the multistage technique, the coefficient of variation was extremely high for the red oak stratum due to its low estimated total volume, as seen in Table 4. Also, the volume per unit area was low as shown in Table 7. On the other hand, the application of the multiphase technique resulted in a high estimated total volume for the red oak stratum, with an estimated standard deviation almost two times as large as that obtained with the multistage sampling method.

Table 4 also shows the results for the multiphase technique. The coefficient of variation for the red oak stratum is very low compared with the one obtained by the multistage sampling technique. This difference in the coefficient of variation is due, of course, to the difference in the estimated total volume. If the estimated total volume given by the multistage technique had been of the same order of magnitude as that for the multiphase technique, the former might have given a lower coefficient of variation.

As seen in Appendix E, the plots measured for the multistage technique show slightly higher volumes than those measured on the multiphase technique. The plots were located in the same stand, with soils appropriate for the red oak. In some areas where plots were measured with site index of 80 (Buchnan, 1985), the factor that determined the differences in volumes per plot was the density. As seen from Tables 5 and 20, the total and merchantable basal area for red oak for the plots measured on the multistage technique are higher than those for the multiphase technique. These differences may be a reflection of the management done in the stand. It also indicates the tendency of the multistage technique to concentrate the sampling units in areas of higher volume.

The approximate 95% confidence interval for the estimated total volume of the red oak stratum, calculated based on the estimations given by the multistage technique (Table 4) is extremely large. This is an indication that the information about the parameter being estimated (total volume) is vague (De Vries, 1986). The low estimation of the total volume associated with a high variance determined the large confidence interval.

The confidence interval calculated from the multiphase technique (Table 4) is much smaller than that given by the multistage method. This can only be said on a comparative basis, however. If considered separately, the confidence interval calculated from the multiphase technique is also quite large. Although large, it nevertheless conveys more information about the total volume than the confidence interval calculated from the multistage technique.

In order to reduce the confidence interval, more samples should have been measured in each of the two sampling techniques evaluated. In the case of the multi-stage technique, more primary, secondary and tertiary units would have to be selected. Since only two primary units were located within the red oak stratum and one of them encompassed most of the area of the stratum, the selection of more units would not have been possible. The alternative in this case, would have been to use smaller primary sampling units.

The relationship between the size variable and the predicted volume for the second stage and estimated volume for the third stage for the red oak stratum, were shown on Figures 2c and 3c. None of the correlations were significant at the 0.05 level. Contrary to what would be expected, the correlation between the two variables shown in Figure 2c is negative. This result is due to: a) the six plots selected from the secondary sampling unit with the highest value for the size variable were located on an area that was thinned (the average volume for these plots was 2,113 cubic feet/acre) and b) the selection probability for the secondary unit in question was high (0.23511). This produced a low value for the expansion factor for the second stage

($1/P_{ij}*t_{ij}$). As a result, the predicted volume for the secondary unit in question was low.

According to information obtained from the regional office of the U.S. Forest Service (Norton, 1987), the original basal area of the red oak stand was 110 square feet/acre. It was thinned to 90 square feet/acre. The average total basal area measured in this study varied between 95 and 120 square feet/acre and the merchantable basal area between 85 and 115 square feet/acre.

The other six plots measured on the red oak stand were located in areas that had not been thinned. These plots were selected from the secondary sampling unit with a smaller value for the size variable. The average volume for these plots was 2,645 cubic feet/acre. Both the total and the merchantable basal area varied between 85 and 160 square feet/acre.

The selection probability for the secondary unit under consideration was also high, but lower than for the secondary unit previously considered. Its value was 0.16319. This created a larger value for the expansion factor for the second stage when compared to the other secondary unit. This fact, associated with the higher volume per plot, determined a larger predicted volume for the secondary unit.

The correlation between percent crown closure and the predicted volume for the third stage is positive, but relatively low (Figure 3c). This low correlation is due to the high variability of the volumes within the measured plots. The regression line shown is not significant at the 0.05 level. There are plots with the same percent crown closure but with quite different volumes. The plots with a crown closure of 85%, for instance, show a great variety of volume. This is an indication of the inadequacy of measuring percent crown closure as a size variable.

Table 13 shows the changes in the estimated total volume and its variance for the red oak stratum as calculated by changing one of the selection probabilities at a time, while keeping the others as originally calculated. As seen in Table 13, only three trials are shown for the probability P_i: trials #1, 2 and 4. Trial #2 presented the same values for the estimated total volume and its variance as the values shown in Table 4, for the multistage sampling technique. This is because the values of the selection probability for trial #2 were equal to the values originally calculated. The same happened to the value of the probability P_i for trial #3.

Table 14 shows that the largest increase in the estimated total volume and its variance, as compared to the values shown in Table 4, occurred with trial #5 for the probability P_{ij}. The values of this probability determined an increase on the estimated total volume of 11.1% and on the variance of 23.5%.

The larger reduction on the estimated total volume and its variance occurred with trial #1 for the probability P_{ij}.

The estimated total volume was reduced by 6.4% and its variance by 13.2%.

Trial #2 for the probability P_{ijk} gave the smallest changes on the estimated total volume and its variance, as compared with the values on Table 4. The estimated total volume was increased by just 1.0% and its variance by only 2.1%.

Figure 6a and 6b shows the effect of combining the selection probabilities on the estimated total volume and its variance. The pattern is identical for both variables. Group A refers to the combinations 1 to 25; group B to the combinations 26 to 50 and group C to the combinations 51 to 75.

The pattern of the curve repeats itself in each group on both graphs. Each line parallel to the X-axis (on both graphs) represents a combination of the selection probabilities. For example, the first small horizontal line to the left, close to the Y-axis, refers to the combination 1-1-1 of the selection probabilities. This means that the values of trial #1 of the probabilities P_i, P_{ij} and P_{ijk} were introduced in equations 5 and 6. Combination 1-1-2 means that the values for trial #1 of probability P_i and P_{ij} and trial #2 for P_{ijk} were introduced in equations 5 and 6. All the 75 combinations that resulted for the red oak stratum, and their effects on the estimated total volume and its variance can be followed by Figure 6a and 6b, respectively, and Table 14.

When the first five combinations were completed and the values for trial #2 for the probability P_{ij} were introduced in equations 5 and 6, a sudden increase in the estimated total volume and its variance occurred (Figure 6a and 6b). This increase occurred at combination 1-2-1. As shown in Table 14, the values of the probability P_{ij} for trial #2 and of P_{ijk} for trial #1 determined the observed increase in the estimated total volume.

The last five combinations of probability on all the groups have the highest values for the estimated total volume and the variance. This is the result of the combined effects of the probabilities P_{ij} and P_{ijk}. As seen in Table 13, trial #5 of the probability P_{ij} determined the highest increase in the estimated total volume and its variance. Trial #5 of the probability P_{ijk} also determined an increase on the estimated total volume and its variance.

The minimum value for the estimated total volume was obtained by the values of the probabilities for the trials 1, 1 and 4, for P_i, P_{ij} and P_{ijk}, respectively. The values for each of the selection probabilities for those trials produced a decrease in the estimated total volume (Table 14). The value for the minimum estimated total volume, shown on Table 15, represents a reduction of 11.0% as compared to the value in Table 4, for the multistage sampling technique.

The maximum value for the estimated total volume was obtained with the values of the probabilities for trials 4, 5 and 5, for P_i, P_{ij} and P_{ijk}, respectively. The values for each of the selection probabilities for those trials (Table 13) determined an increase in the estimated total volume. The maximum value for the estimated total volume, shown on Table 15, represents an increase of 20.0% in relation to the value in Table 4.

The minimum and maximum values for the estimated variance of the total were also obtained by the same combinations of probabilities as for the volume. The minimum value of the estimated variance of the total, shown in Table 15, represents a reduction of 20% in relation to the value in Table 4. The maximum value represents an increase of 39% in relation to the value in Table 4.

5.2 **Conifer Stratum** For the multistage technique evaluated, the conifer stratum was not subdivided into jack pine and red pine as was the case for the multiphase technique. The basic reason for not subdividing was that the jack pine plantations in the study area represent only 7.3% of the area of the conifer stratum. As a result, to evaluate the jack pine as a separate stratum by the multistage technique, a different structure would be required in terms of the number of stages and the sizes for the sampling units at each stage. For the multiphase technique, however, the subdivision of a stratum is a requirement of the procedure.

Comparing the results obtained by the multistage and the multiphase sampling techniques presented in Table 4 for the conifer stratum, it can be seen that the multistage technique gave an estimated total volume 13% higher than that estimated by the multiphase technique. The estimated standard deviation from the multistage method was 3 times larger than that for the multiphase approach. As a result, the coefficient of variation for the multistage sampling technique was larger than the coefficient of variation for the multiphase technique.

The approximate 95% confidence interval calculated for the multistage technique is much larger than the one calculated for the multiphase method. This is due to the larger standard deviation of the former technique as compared to the later.

The results obtained from the multistage technique reflect mostly the volume of red pine. Of the twelve plots measured, only one had a predominance of jack pine. Four plots contained only a few jack pine trees and seven plots were composed of only red pine.

The results obtained from the multiphase technique can be considered as a more realistic appraisal of total volume on the conifer stratum. Since the stratum had to be subdivided into red pine and jack pine, plots were measured on both substrata.

Table 19 presents the results obtained for each substratum. As shown, the red pine substratum presents a higher volume per unit area and a higher standard deviation than the jack pine substratum. The higher standard deviation may be associated with the management that is being done on the red pine stands. The jack pine stands, on the other hand, are not being managed. They are relatively old and may have reached a stage of nearly homogeneous volume.

The above discussion explains why the multistage technique presented a higher value for the estimated total volume. Most of the plots measured had a predominance of red pine trees or were pure red pine stands. These stands have higher values for the volume per unit area and for the basal area (see Table 20) than the jack pine stands.

In the multiphase technique, the estimated total volume for the conifer stratum had to be averaged between the plots measured on the red pine and on the jack pine substrata.

The pine plantations are concentrated mostly in the southeast portion of the study area, although there are some small-area plantations in the northern part. The two primary sampling units selected within the conifer stratum for the multistage technique were the same. In other words, only one primary unit was selected twice. As the sampling is done with replacement, the four secondary units were also selected within the same primary unit. Each time a unit is selected, a new selection of units on the subsequent stages has to be done (sampling with replacement).

Figure 2b shows the relationship between the size variable and the predicted volume for the second stage sampling of the conifer stratum. Each point on the graph represents a secondary sampling unit. The two points aligned on the X-coordinate at 70 acres correspond to the same secondary unit which was selected twice. As was the case of the red oak stratum, the conifer stratum exhibits a high negative correlation between the two variables.

The secondary unit that contained the greatest estimated area of the conifer stratum had the lowest predicted volume. Among the secondary units selected, this particular unit had the highest selection probability (0.18470) which determined a low value for the expansion factor ($1/P_{ij}*t_{ij}$) for the unit. As a result, the predicted volume for the secondary unit in question was low even though the three plots measured in this secondary unit had an average volume of 2,544 cubic feet/acre. These plots were all located in pure red pine stands with merchantable basal areas between 90 and 130 square feet/acre.

The secondary unit that was selected twice also had quite different values for its estimated volume, as seen on Figure 2b. The three plots chosen the first time the secondary unit was selected had higher estimated volumes on average 2,966 cubic feet/acre with merchantable basal area varying between 110 and 180 square feet/acre. Two of these plots were in pure red pine stands. The third was measured in a mixed stand of red pine and jack pine. On the other hand, the three plots that were selected the second time the secondary unit was chosen had an average volume of 1,918 cubic feet/acre, with merchantable basal area varying between 70 and 130 square feet/acre. Since the expansion factor value for this twice-chosen secondary unit was high, the reason for the difference in predicted volume was simply the differences in the measured volumes of the plots.

The secondary unit with the smallest area has the second highest predicted volume. All three plots measured within this secondary unit were located in pure red pine stands. Their average volume was 2,342 cubic feet/acre with merchantable basal area varying between 90 and 130 square feet/acre. The value of the expansion factor for this secondary unit was the highest among the four, since its selection probability was the lowest. This, in conjunction with the location of the field plots, determined its high predicted volume.

Figure 3b shows the relationship between the size variable and the estimated volume for the third stage samples. The correlation is positive and relatively low (r=0.54). This is slightly greater than the correlation achieved for the red oak stratum (r=0.512) shown on Figure 3c. As was the case for the red oak stratum, there are several plots in the conifer stratum with the same value for the percent crown closure, but with different volumes. The low correlation is, of course, associated with the variability of the estimated plot volumes. The inadequacy of the percent crown closure measurement as a size variable for calculating the selection probabilities is again demonstrated. Both regression lines for the conifer stratum were not significant at the 0.05 level.

Table 9 presents the estimated total volumes and their variances that resulted from changing the selection probabilities for each stage of the multistage technique. Each selection probability was changed in turn, while keeping the others as originally calculated. As shown in Table 10, the highest increase in both estimated total volume and its variance was observed with trial #4 of the probability P_{ijk}. The increase in volume was 11% coupled with a 24% increase in variance as compared with the values shown in Table 4, for the multistage sampling technique.

The largest decrease observed in the estimated total volume, compared with the values in Table 4, was observed with trial #1 for the probability P_i. Under the conditions of this trial, the volume was reduced 15%. The largest decrease in the estimated variance of the total, on the other hand, occurred with trial #5 for the probability P_{ij} where a reduction of 22% was observed.

The least variation in the estimated total volume and its variance, in comparison with the results in Table 4, were observed with trial #2 for the probability P_{ij}. The volume was increased by only 0.8% and the variance by only 1.6%. As shown in Appendix B, the values of the probability P_{ij} for trial #2 are very close to the ones originally calculated.

Figure 5 shows the effect of changing the selection probabilities on the estimated total volume and its variance. Group A refers to the combinations of the selection probabilities 1 to 25; group B to the combinations 26 to 50; group C to the combinations 51 to 75; group D to the combinations 76 to 100 and group E to the combinations 101 to 125.

The pattern for the volume curve (Figure 5a), was basically determined by the changes in the probability P_{ijk}. The 125 combinations and their effects on the estimated total volume can be followed by Figure 5a and Table 10. Once again, each of the small horizontal lines on the graph represent a combination of probabilities.

The first combination of the selection probabilities (1-1-1) determined a certain value for the estimated total volume. The second combination (1-1-2) determined a smaller value for the estimated total volume than the first. Although the probability P_{ij} for trial #1 and P_{ijk} for trial #2 had the effect of increasing the estimated total volume, the decreasing effect of the probability P_i was dominant. The same happened for the third combination (1-1-3).

When the values for trial #4 of the probability P_{ijk} were introduced in equation 5, a noticiable increase in the estimated total volume occured. This was due, in part, because trial #4 for the probability P_{ijk} had a strong effect on increasing the estimated total volume. This, in combination with the increasing effect of the values of trial #1 for the probability P_{ij}, surpassed the decreasing effect of trial #1 for the probability P_i.

The introduction of the values for trial #5 of the probability P_{ijk} in equation 5 resulted in a marked decrease on the estimated total volume observed on combination 1-1-5. As seen on Table 10, the values for trial #5 of the probability P_{ijk} had the effect of decreasing the estimated total volume. This effect, in combination with the decreasing effect of trial #1 for the probability P_i, were responsible for the decreased estimated total volume.

The decreasing trend of the curve within group A is due to the outcomes of trials #4 and 5 for the probability P_{ij}. As seen in Table 10, the values of the probability P_{ij} on those trials resulted in a continual decline of the estimated total volume.

The sudden increase in the estimated total volume that occurred at the beginning of group B was due to probability P_i. At that point the values of trial #2 for the probability P_i were introduced into equation 5 (combination 2-1-1). As shown in Table 10, the effect of increasing the estimated total volume for all three probabilities resulted in the sharp increase.

The values of the first three trials of the probability P_{ijk} had the effect of increasing the estimated total volume, as seen on the last line of Table 10. It should be noted, though, that the percentage of increase varies from trial to trial. Actually it decreases from 5.9% to 2.0%. This aspect determined the step-like pattern for the first three combinations which repeats itself at constant intervals, as can be seen on Figure 5a. As a result, even if the combination of the probabilities has a tendency to increase the estimated total volume, as occurred with the combinations 2-1-1 to 2-1-3, the pattern of the curve remains because of the different percentages of increase.

The pronounced increase that occurred at the beginning of group B was not repeated at the beginning of group C. At the beginning of group C, the values for trial #3 of the probability P_i were introduced into equation 5. As seen in Table 10, those values decreased the estimated total volume. As a result, the pattern of the curve remained in a decreasing tendency. Another abrupt increase occurred at the beginning of group D, because of the introduction of the values of trial #4 for the probability P_i.

The discussion above is also valid for the curve of the variance of the total, shown on Figure 5b. The pattern of this curve is similar to the pattern of the volume curve (Figure 5a).

The minimum values for the estimated total volume and its variance were obtained with the trials #1, 5 and 5 for P_i, P_{ij} and P_{ijk}, respectively. As seen on Table 10, the percentage of decrease on both the estimated total volume and its variance are the largest for each probability. The minimum value for the estimated total volume, shown on Table 15, corresponds to a reduction of 29% in relation to the value on Table 4. The variance decreased by 45.0% in relation to the value on Table 4, for the multistage sampling technique.

The maximum value for the estimated total volume and its variance were obtained with trial #2, 3 and 4 for P_i, P_{ij} and P_{ijk}, respectively. These trials correspond to the highest percentages of increasing both the estimated total volume and its variance, for each probability (Table 10). The maximum value for the estimated total volume, shown on Table 15, corresponds to an increase of 25.0% in relation to the value on Table 4. The increase in variance, corresponds to 43% in relation to the value on Table 4.

5.3 **Hardwood Stratum** Comparing the results obtained for the multistage and multiphase sampling techniques presented in Table 4 for the hardwood stratum, it can be seen that multistage technique gave a much higher estimated total volume and standard deviation than the multiphase technique. The estimated total volume for the multistage technique is 2.5 times higher than the multiphase technique and the estimated standard deviation is 1.7 times larger. The coefficient of variation for the multiphase technique on the other hand, is much higher than that for the multistage due to the lower estimated total volume. This lower estimated total volume produced in turn an extremely large confidence interval. The confidence interval estimated by the multistage technique is also large but not as large as for the other method. In both cases, the information provided by the confidence intervals about the parameter being estimated is vague.

The multistage technique showed its tendency to sample areas of higher volumes. The plots measured for this technique had an average volume of 2,202 cubic feet/acre. The average volume of the plots measured in the multiphase technique, on the other hand, was 1,957 cubic feet/acre. The lower average merchantable basal area of the plots measured in the multiphase technique as compared to the multistage technique (Tables 20 and 5 respectively) reflects the lower average volume per plot and also the lower estimated total volume.

Figure 2a shows the relationship between the size variable and the predicted volume for the second stage of the hardwood stratum. Each point on the graph represents a secondary unit selected. As shown there is no secondary unit that was selected twice as occurred on the other two strata. The correlation between the two variables is negative and much lower than correlation for the other two strata. The secondary unit with the highest predicted volume contained the smallest estimated area of hardwood, whereas the secondary unit with the lowest predicted volume contained the largest estimated area of hardwood. This negative correlation between the two variables is contrary to what would be expected.

The secondary unit that contained the largest area of the hardwood stratum was the one that had a large value for its selection probability (0.14945). This high selection probability determined a low value for the expansion factor for the second stage ($1/P_{ij}*t_{ij}$). As a result, the predicted volume for the secondary unit in question was low, even though the three plots measured on this secondary unit had an average volume of 2,071 cubic feet/acre. The merchantable basal area of these three plots varied between 100 and 115 square feet/acre.

At the other extreme, the secondary unit that contained the smallest area of the hardwood stratum had one of the largest predicted volumes because of its low selection probability (0.06253). This low value of the selection probability determined a high value for the expansion factor ($1/P_{ij}*t_{ij}$) which in term determined the higher volume for the secondary unit in question. The average volume for the three plots measured on the secondary unit in question was 2,107 cubic feet/acre. The merchantable basal area of these three plots varied between 100 and 115 square feet/acre.

Figure 3a shows the relationship between the size variable and the estimated volume for the third stage. The correlation is positive but extremely low. As in the case of the other two strata, there are several plots with the same percent crown closure but with quite different volumes. The low correlation is associated with the high variability among the plot volumes. The inadequacy of the size variable used is evident. The high variability of the volume in the three strata is clear from Figure 3. As with the conifer and red oak strata, both correlations for the hardwood stratum were not significant at the 0.05 level.

Table 11 presents the effects of changing the selection probabilities on the estimated total volume and its variance, for the hardwood stratum. The calculations were performed as for the other two strata, by changing one probability at a time and keeping the others as originally calculated.

The highest increase in both estimated total volume and its variance was obtained with trial #5 for the probability P_{ijk} (Table 12). The total volume was increased by 13% and the variance by 28% as compared to actual sample values for the multistage sampling technique (Table 4). The largest decrease in the estimated total volume occurred with trial #4 for the probability P_i when the volume was reduced 9.7%. The largest decrease in the estimated variance of the total occurred with trial #5 for the probability P_{ij}. The reduction in the variance was of 17.7%.

The least variations in the estimated total volume and its variance occurred with trial #4 for the probability P_{ijk}. The volume was increased by only 0.7% and the variance by 1.5%. Another small variation on both estimates was observed with trial #4 for the probability P_{ij}. The estimated total volume was increased by 0.9% and its variance by 1.8%. As seen in Appendix B, the values of selection probabilities for trial #4 are close to the original probabilities.

Figure 4a and b shows the effect of changing the selection probabilities on the estimated total volume and its variance, respectively. Group A refers to the combinations of the selection probabilities 1 to 25; group B to the combinations 26 to 50; group C from 51 to 75; group D from 76 to 100 and group E from 101 to 125.

The pattern of the curve shown on Figure 4a and b for the first five combinations was determined by the probability P_{ijk}. As shown in Table 12, the probability P_{ijk} had, in all of the trials, an effect of increasing both the estimated total volume and its variance. It is noticiable that the percentage of increase varies from trial to trial. There is an oscillating character to the variation. The sudden increase in the estimated total volume and its variance occurred when the values of the probability P_{ijk} for trial #5 were introduced into equations 5 and 6, respectively. As seen on Table 12, the values of the probability P_{ijk} for trial #5 gave highest rate of increase on both the estimated total volume and its variance. The pattern generated by the probability P_{ijk} repeats itself through the curves.

The other marked increases and decreases are due to the introduction of the other two probabilities (P_{ij} and P_i) into the equations to estimate the volume and its variance. The start of each group of combinations is done by changing the probability P_i. As a result, just the probability P_{ij} and P_{ijk} were changed within each group.

As seen on both curves, the values for the estimated total volume and its variance for group B and E are overall greater than those for the other three groups. These differences are due to the probability P_i. As shown in Table 12, the effect of the values of the probability P_i for trials #2 and 5 is to increase the estimated total volume and its variance. These trials correspond respectively to the beginning of groups B and E. The pattern of both curves within these two groups are similar to the other groups and were determined, as stated, by the probabilities P_{ij} and P_{ijk}.

The maximum value for the estimated total volume and its variance did not occur, as happened on the other strata, at the same combinations of probabilities. The maximum value for the estimated total volume occurred at trials #5, 3 and 5 for P_i, P_{ij} and P_{ijk}, respectively. As seen in Table 12, the values for the three probabilities at those trials gave the highest rate of increase on the estimated total volume. The maximum value of the estimated total volume shown on Table 15, corresponds to an increase of 25.8% in relation to the value on Table 4, for the multistage sampling technique.

The maximum value for the estimated variance of the total, on the other hand, occurred at trials #2, 3 and 5 for P_i, P_{ij} and P_{ijk}, respectively. The values for the three probabilities at those trials gave the highest rate of increase on the estimated variance (Table 12). The corresponding maximum value for the estimated variance, shown in Table 15, represent an increase of 49.2% in relation to the value on Table 4.

The minimum values for both the estimated total volume and its variance, occurred at the same combination of the probabilities trials #4, 5 and 4 for P_i, P_{ij} and P_{ijk}, respectively. As seen in Table 12, for the probability P_i, trial #4 gave the highest reduction of both the estimated total volume and its variance. For the probability P_{ij}, the same happened for trial #5. For the probability P_{ijk}, all the trials had the effect of increasing the estimated total volume and its variance, but the rate of increase was the lowest.

The value of the minimum estimated total volume shown in Table 15 represents a reduction of 18.1% in relation to the value in Table 4. For the variance, the minimum value represents a reduction of 26.9% in relation to the value in Table 4, for the multistage sampling technique.

5.4 **General Considerations** From the discussions related to Figure 2a, b and c, one may get the impression that negative correlation a size variable and predicted volume always occurs with the multistage technique. There is a general explanation for the common behavior observed on the correlation between the size variable and the predicted volume for the second stage. The size of the primary units caused, in most cases, a high variability for the estimated area of the stratum contained on the 16 secondary units. The places where the field plots were measured also influenced the behavior of the curves. These effects were discussed separately for each stratum.

For the red oak stratum, of the 16 secondary units within the primary unit selected twice, four had a zero value for the estimated area of the stratum. There are two possible reasons for the zero value: the secondary unit was located within the limits of private land, or was located outside the boundaries of the stratum. Two secondary units had very low values for the estimated area of the stratum.

For the conifer stratum, of the 16 secondary units within the primary unit selected twice, seven had an area of zero and one had a very low value for the estimated area of the stratum. The reasons for the zero area estimates were the same as above.

As a result of the zero area estimates and of the low estimated area on several of the secondary units on the above strata, the total sum of the areas of the 16 secondary units was relatively low. This resulted in very high selection probabilities for secondary units with high estimated area of the stratum, and very low selection probabilities for units with low estimated area of the stratum. These selection probabilities are used to calculate the expansion factor for the second stage ($1/P_{ij}*t_{ij}$).

If, within a primary unit, a secondary unit is selected which has a high selection probability (large estimated area of the stratum) and another secondary unit is selected which has a low selection probability (small estimated area of the stratum), the expansion factor for these two units would be quite different. The first secondary unit selected would have a low expansion factor because of its higher selection probability. The second secondary unit would have a high expansion factor because of its lower selection probability. The end result would be that the secondary unit with the largest estimated area of the stratum would have a lower predicted volume than the unit with the smaller estimated area of the stratum. This would happen only if there were not large differences on the volumes of the plots measured within each secondary unit.

Two primary units were selected in the hardwood stratum. The first was completely contained within government property. As a result, none of the secondary units had a zero value for the estimated area of the stratum. Only one secondary unit had a low estimated area of the stratum. Accordingly, all the selection probabilities calculated for the 16 secondary units were smaller than 0.1, and they did not vary as much as in the other strata. On the other two strata, some of the selection probabilities were much higher than 0.1 (e.g., in the red oak stratum).

For the two secondary units selected within the first hardwood-stratum primary unit, the proportionality between the estimated area of the stratum and the predicted volume was observed (Figure 2a). The two points with the highest volumes correspond to the two secondary units selected within the first primary unit. The secondary unit with small estimated area of the stratum also had a small predicted volume, and vice-versa. This proportionality was observed because of the lower variability found among the calculated selection probabilities. It should be noted that the proportionality between the size variable and the predicted volume has to occur among secondary units within a single primary unit. The proportionality between the size variable and the predicted volume among secondary units of different primary units may occur independently of the variation existing on the selection probabilities.

For the two secondary units chosen within the second primary unit, the proportionality between the estimated area of the stratum and the predicted volume, although not very evident, can also be observed. The two points located on the lower portion of Figure 2a correspond to the two secondary units selected within the second primary unit. Of the 16 secondary units contained in the second primary unit, one had zero area and four had a relatively low estimated area of the stratum. This produced a large variation among the selection probabilities. Several probabilities, including those for the selected secondary units, had values above 0.1. The consequence was that the expansion factor for these secondary units were the lowest among the four. As a result, these secondary units had the lowest predicted volumes although they had large estimated areas.

One conclusion from the above discussion is that one has to be alert when choosing the size of the primary units in situations where stratification is necessary. The stratum with the smallest area should be used as a guide in order to define the most appropriate size for the primary unit. If this is done carefully, the problems that result from the size variables not being proportional to the predicted volume for a stage may be avoided and more precise estimations may be obtained.

From the discussions related to the effect of changing the selection probabilities on the estimated total volume and its variance, there is no consistent effect. It should be noted, however, that if the change in the probability for the first stage (P_i) determined an increase or decrease on the estimated total volume of x%, the variance was increased or decreased by approximately the same amount. On the other hand, if the change in the probability for the second and third stages, P_{ij} and P_{ijk}, determined an increase or decrease in the estimate total volume of x%, the variance would be increased by approximately 2x%. This aspect is an indication that the size variables used to calculate the selection probabilities for the second and third stages should be measured as accurately as possible.

Of course in practice, no one would know if the volume or the variance is being over or under estimated since the parameters are not known. But one should be aware of the effect that an error on the measurement of the size variable might cause on the estimated total volume and its variance.

5.5 Population Estimates Comparing the results in Table 4 it can be seen that the multistage technique has produced a much higher estimate of the total volume than the multiphase method. Actually, the difference is almost a factor of five times.

According to Langley (1975), the multistage technique can be used in situations where stratification of the population is necessary. In such cases, the method is applied independently on each stratum. The estimated total volumes for the population would be given by the summation of the estimated volume for each stratum. As the estimated volume on each stratum is independent, the variance of the estimated population total is given by the sum of the estimated variance for each stratum (De Vries, 1986). The multiphase technique, on the other hand, is based on stratified random sampling, the difference being that a two-phase, rather than a simple random sample is taken in each stratum (Johnston, 1982).

The major contributor to the higher estimated total volume on the multistage technique is the estimated volume for the hardwood stratum. As seen from Table 4, the estimated total volume for the hardwood stratum calculated by the multistage method is 2.5 times larger than the value calculated by the multiphase technique.

The estimated total volume for the conifer stratum reflects the availability of red pine, which has a higher average volume per acre than jack pine. This also contributed to the larger estimated population total given by the multistage technique. The multiphase estimate of the total volume for the conifer stratum was composed of plots measured on both conifer types. Since jack pine has a lower average volume per acre than red pine, this lowered the average estimated volume for the conifer stratum and, consequently, its estimated total volume.

Had the estimated total volume for the red oak stratum calculated from the multistage technique been of the same order of magnitude as the value for the multiphase technique, the difference between the estimated population total for the methods would have been larger.

Another aspect to consider is the fact that the estimated population total for the multistage technique is simply the summation of the estimated total for each stratum. In the multiphase technique, on the other hand, the estimate of the population total is obtained by multiplying the results from equation 18 by the estimated total area. It is not just a simple addition of the estimated volume per stratum. Equation 18 has the term $w_i = N_i/N$ - called stratum weight - which is a reduction factor, to compensate by the size of the stratum.

In reality, as long as the parameter of the population is not known, it is difficult to say if the estimated population total obtained by the multistage technique is overestimating the population volume or if the estimation of the multiphase is underestimating. But what can be said is that, although smaller, the estimated population total is more precise for the multiphase technique than for the multistage technique. The negative correlation between the size variable and the predicted volume for the second stage and the low positive correlation in the third stage certainly had an effect on the higher estimated population variance. This higher variance determined the lower precision for the estimate of the total volume given by the multistage technique. As a result, in forest inventories in areas of similar forest types, the multiphase technique should be given a first consideration in the process of selecting a multilevel sampling technique.

The results obtained by both sampling techniques evaluated can not be compared with those from the U.S. Forest Service. Forest surveys statistics are accurate at county level. The study area was composed of two townships, as a result, this comparison could not be recommended.

5.6 **Digital Image Processing** The supervised classification of digital imagery implies the selection of training sites. These are "representative" areas of the various cover types of interest. In the present study, training sites were obtained from the forest types of interest (red and jack pines, hardwood and red oak), agricultural areas and water. These training sites were composed of a sample of pixels whose brightness values are the numerical representation of the spectral attributes for each feature type of interest. The classification is ultimately based on the statistics obtained from the various training sites selected.

It is of interest for the image analyst to evaluate the spectral "separability" of the various cover types selected, using different combinations of spectral bands. This evaluation, which helps select a set of bands that best discriminate the classes, can be done statistically or by graphical means (Jensen, 1986). Only the graphical method, restricted to two-dimensional plots, will be presented because of its simplicity and ease of understanding. These graphs are normally called feature space. For details on the use of the statistical procedure, the reader is referred to the above citation and to Goodenough _et al_, 1974.

Figures 7 through 13 show the feature space for all possible two-band combinations of the four MSS LANDSAT bands, and for the first and second principal components. The brightness values used to construct the feature space plots were obtained by a systematic sampling of pixels within each of the training sites selected on the study area. The number of pixels selected was proportional to the size of each training site.

The feature space plot may allow the interpreter to obtain valuable insights into the structure of the multispectral data set. It may also provide an indication of incorrectly chosen pixels (Donker _et al_, 1977).

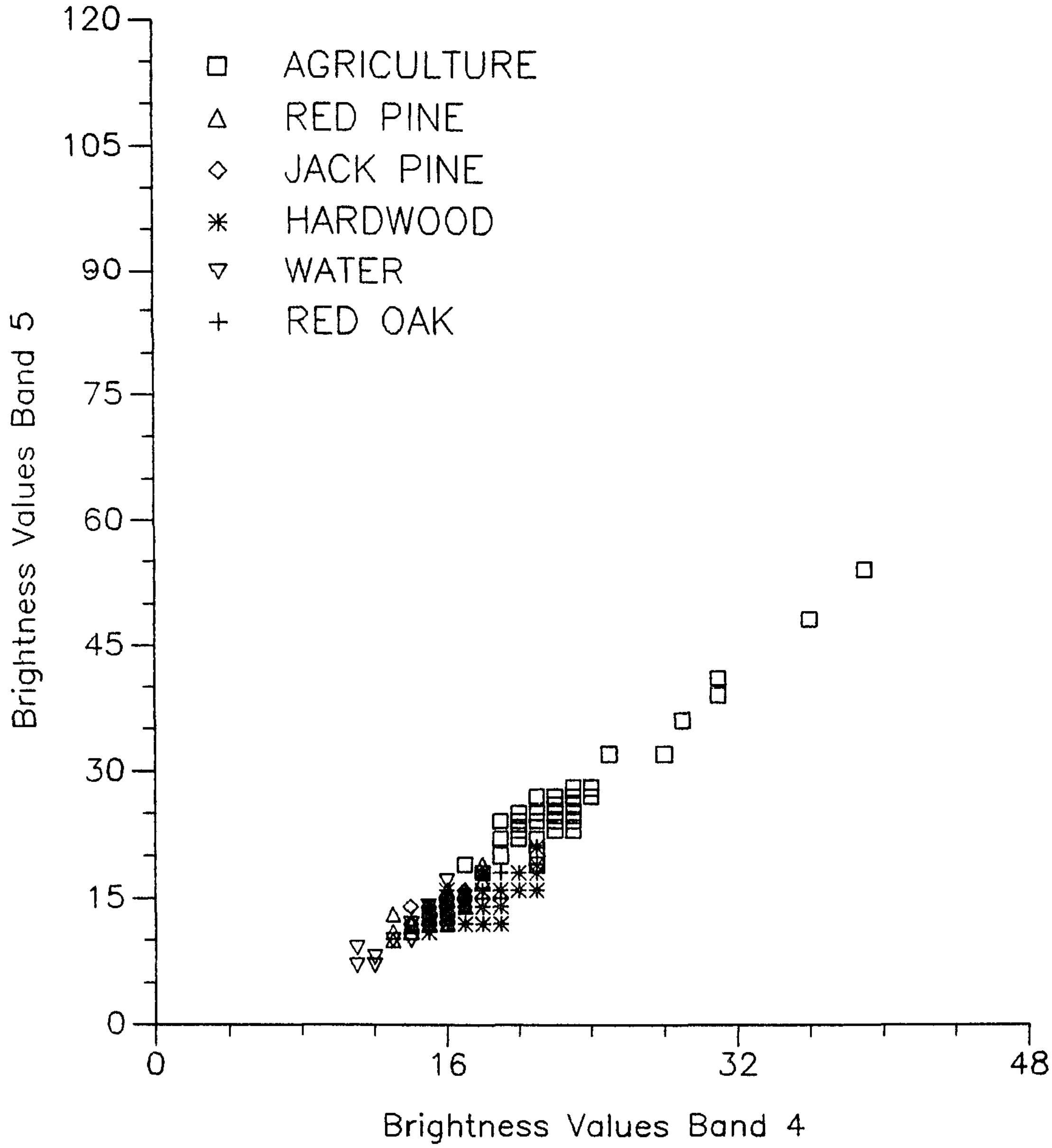


Figure 7 Feature Space for Bands 4 and 5

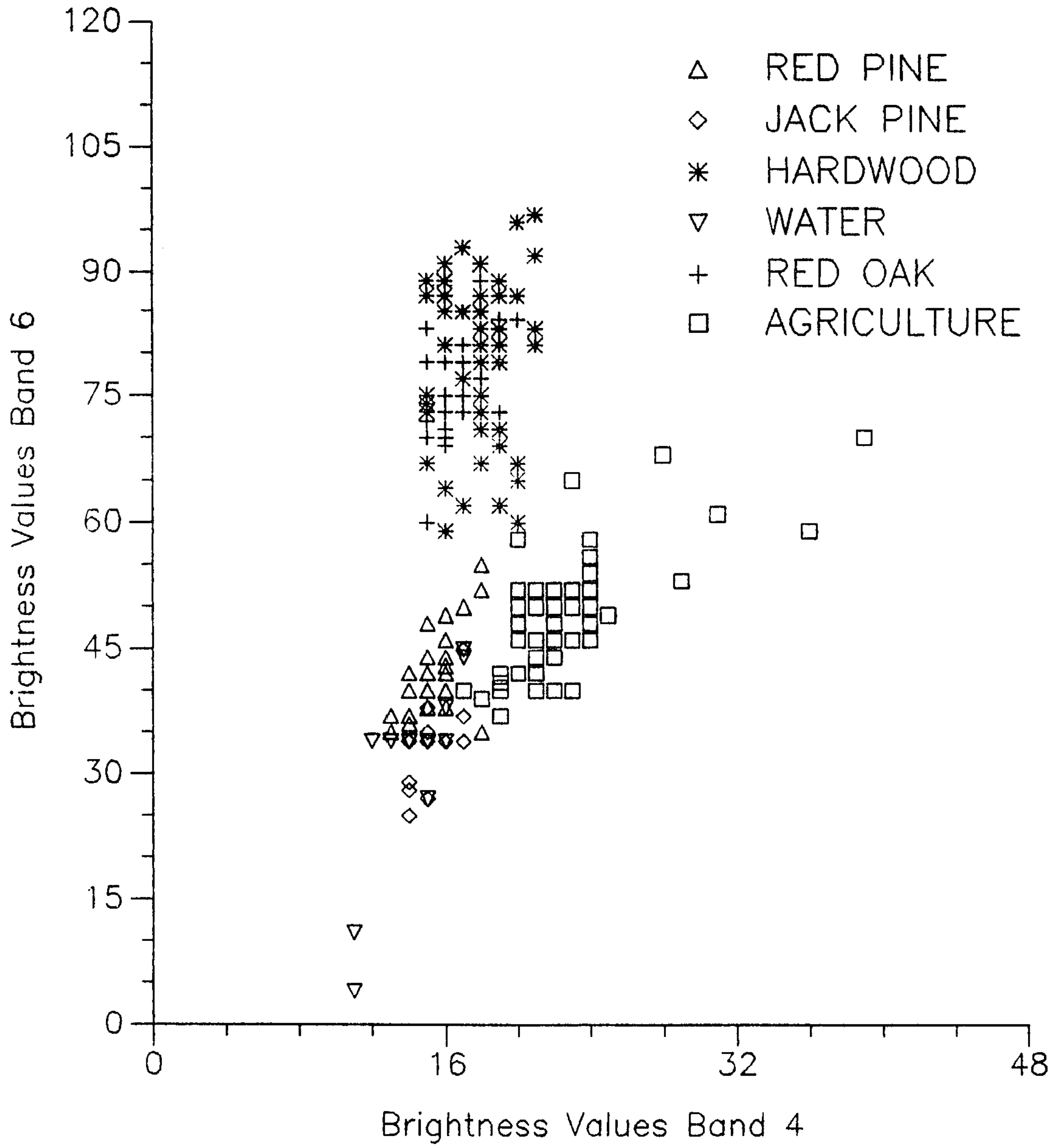


Figure 8 Feature Space for Bands 4 and 6

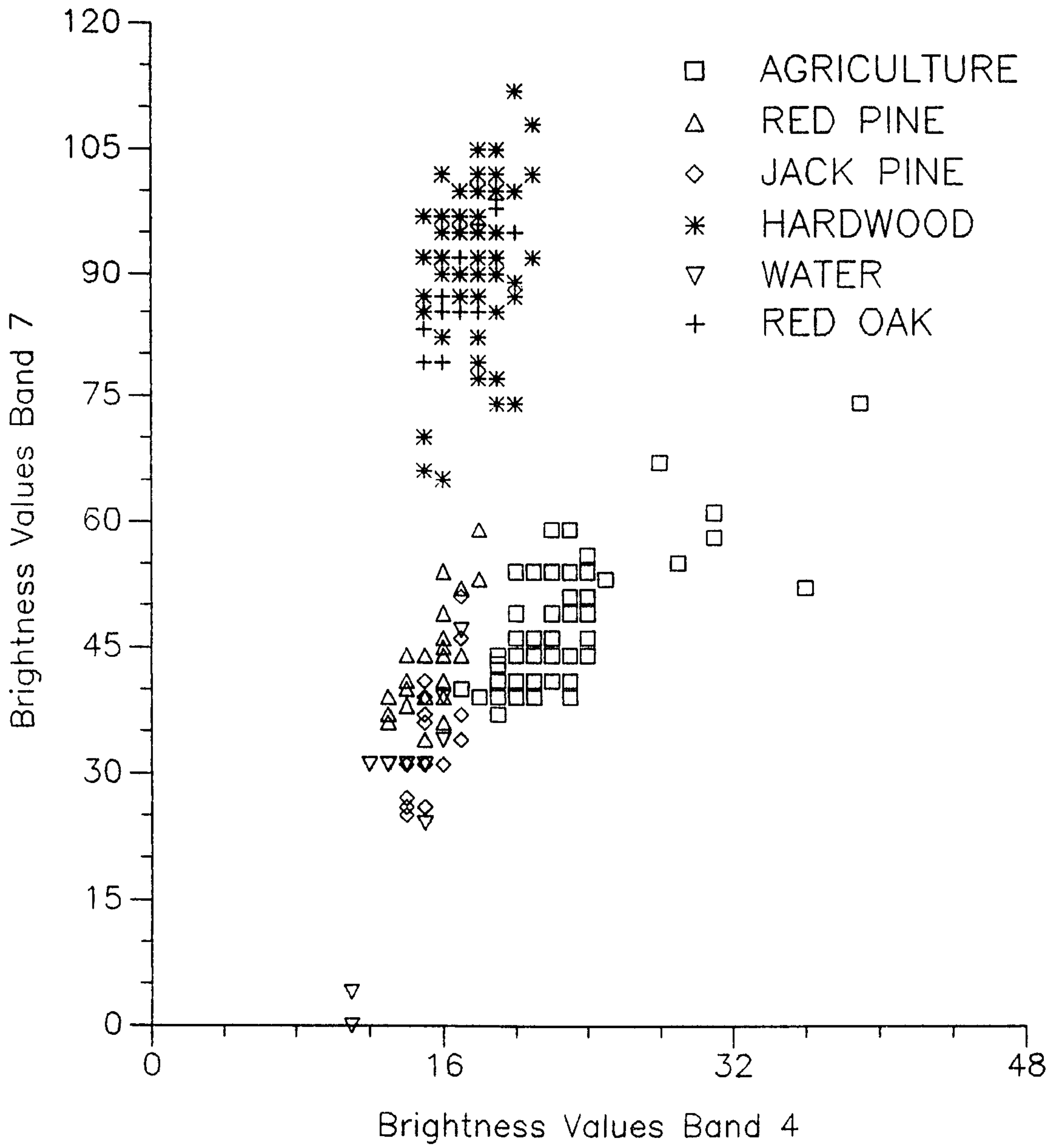


Figure 9 Feature Space for Bands 4 and 7

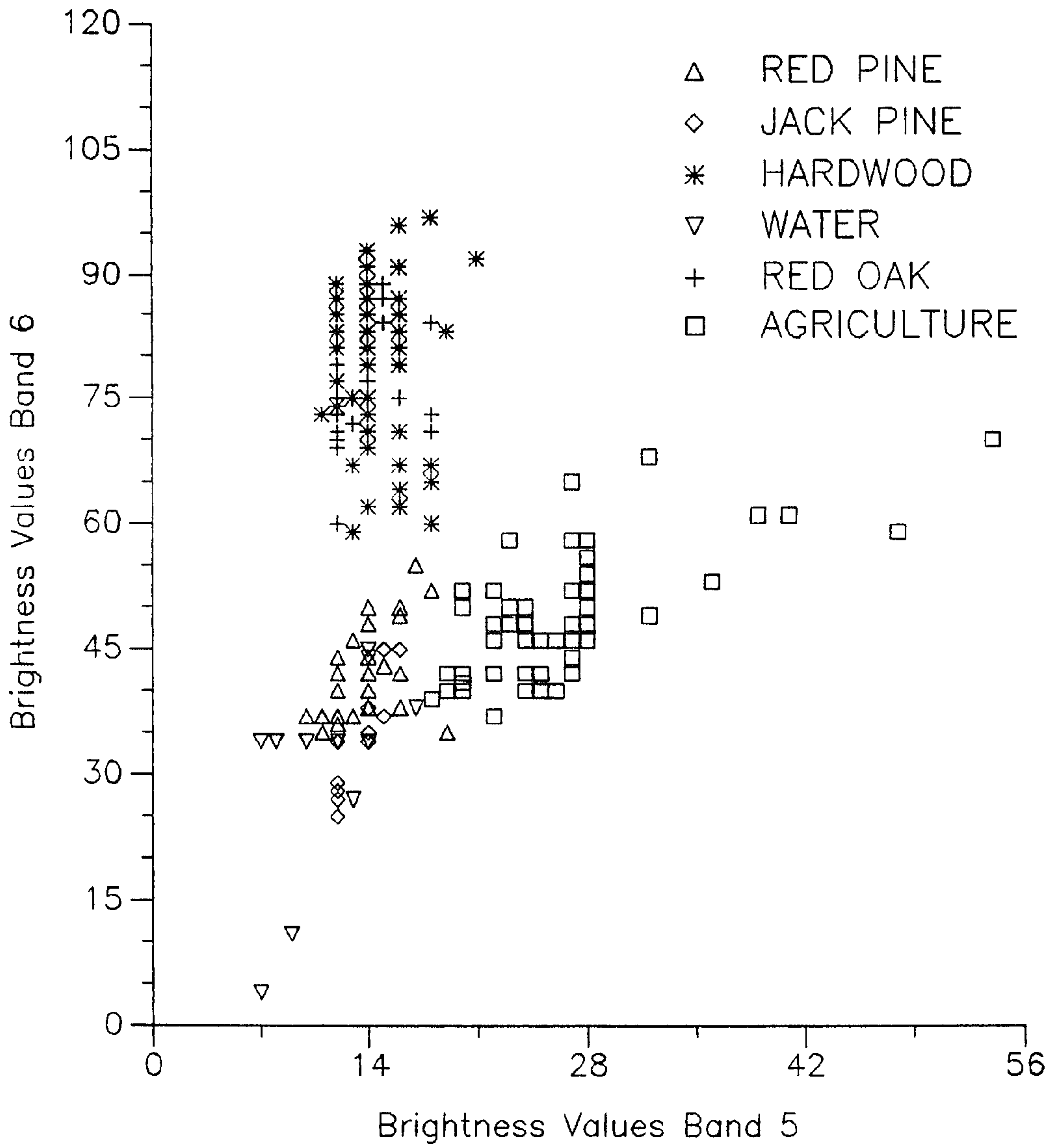


Figure 10 Feature Space for Bands 5 and 6

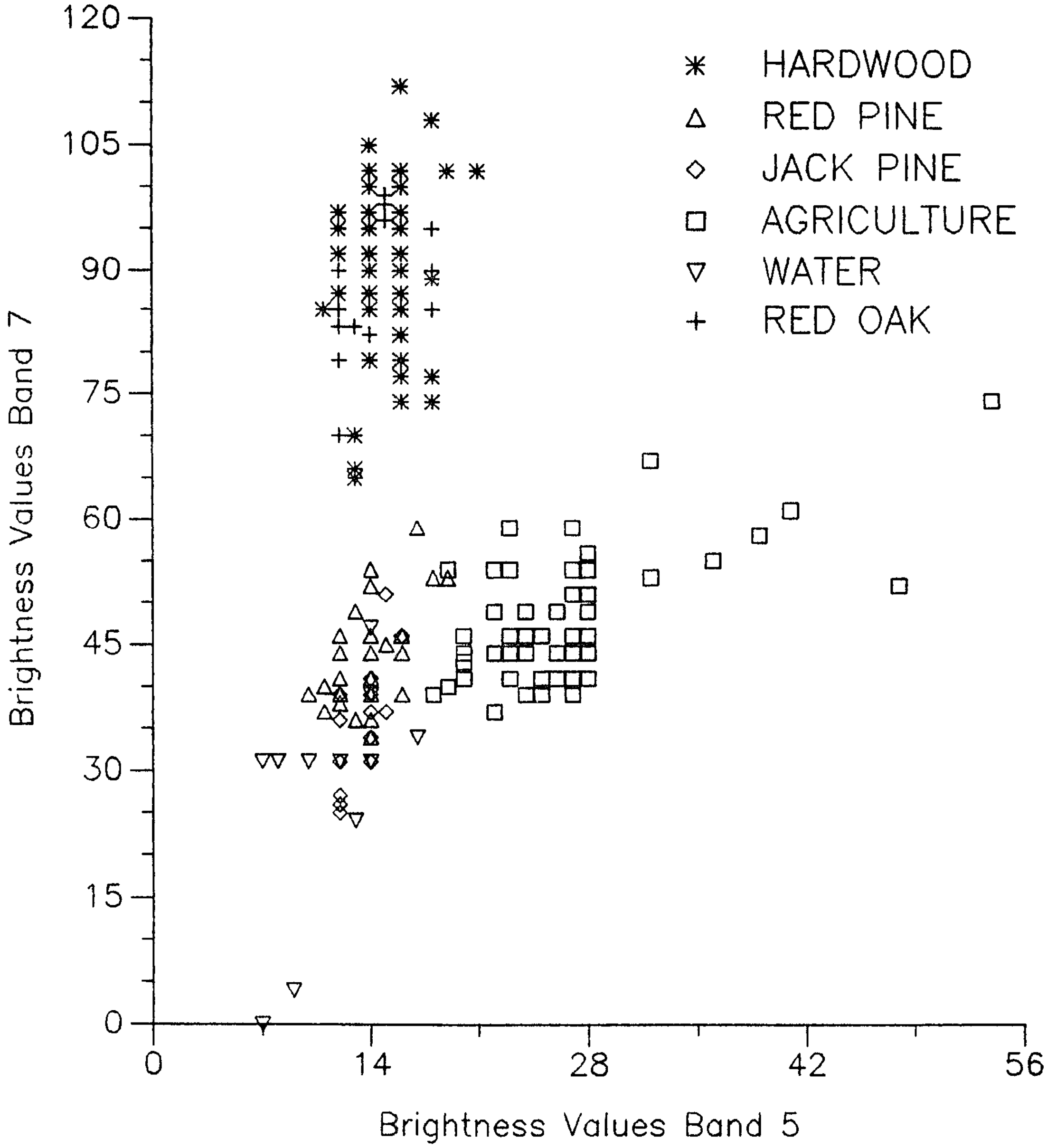


Figure 11 Feature Space for Bands 5 and 7

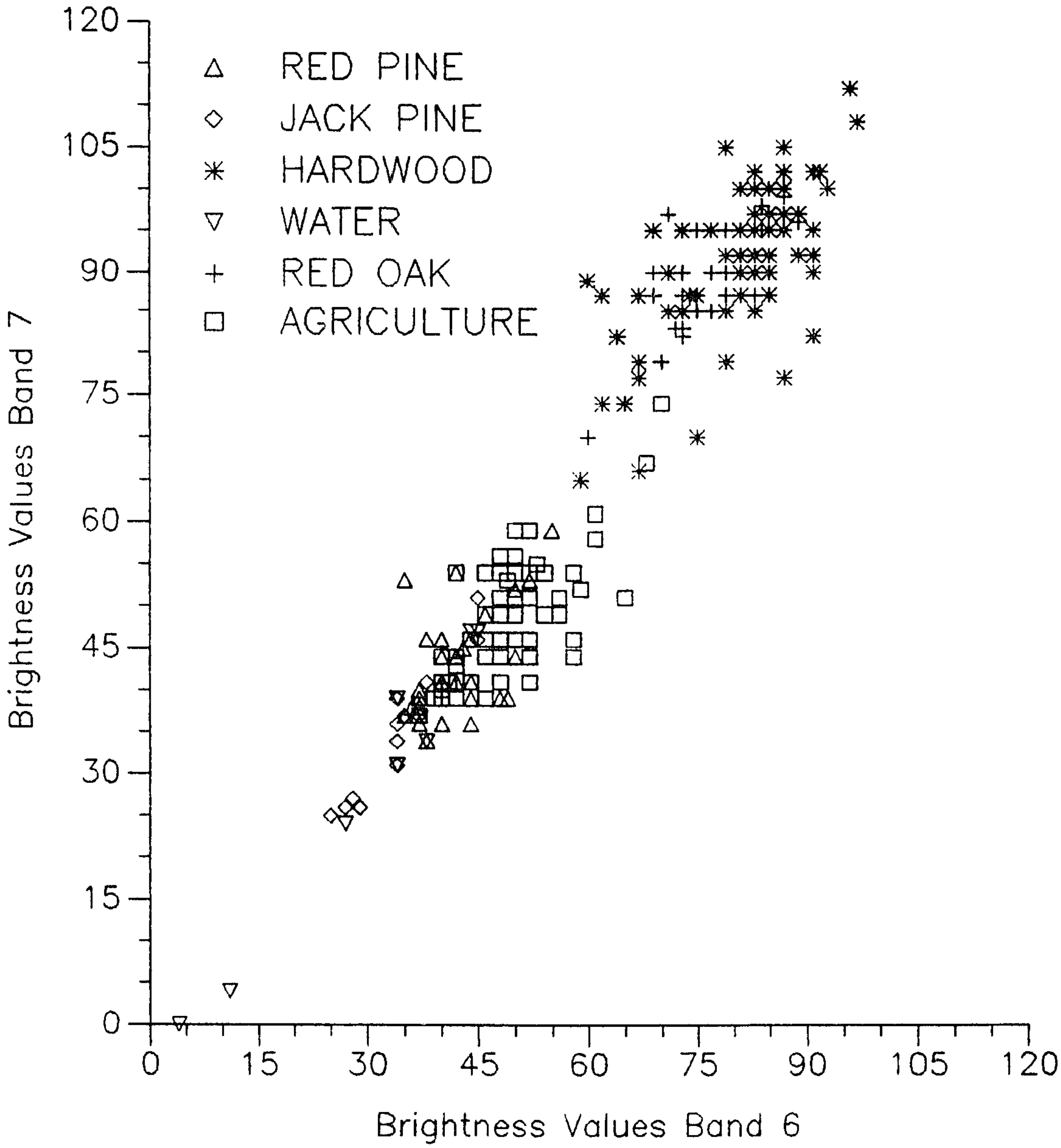


Figure 12 Feature Space for Bands 6 and 7

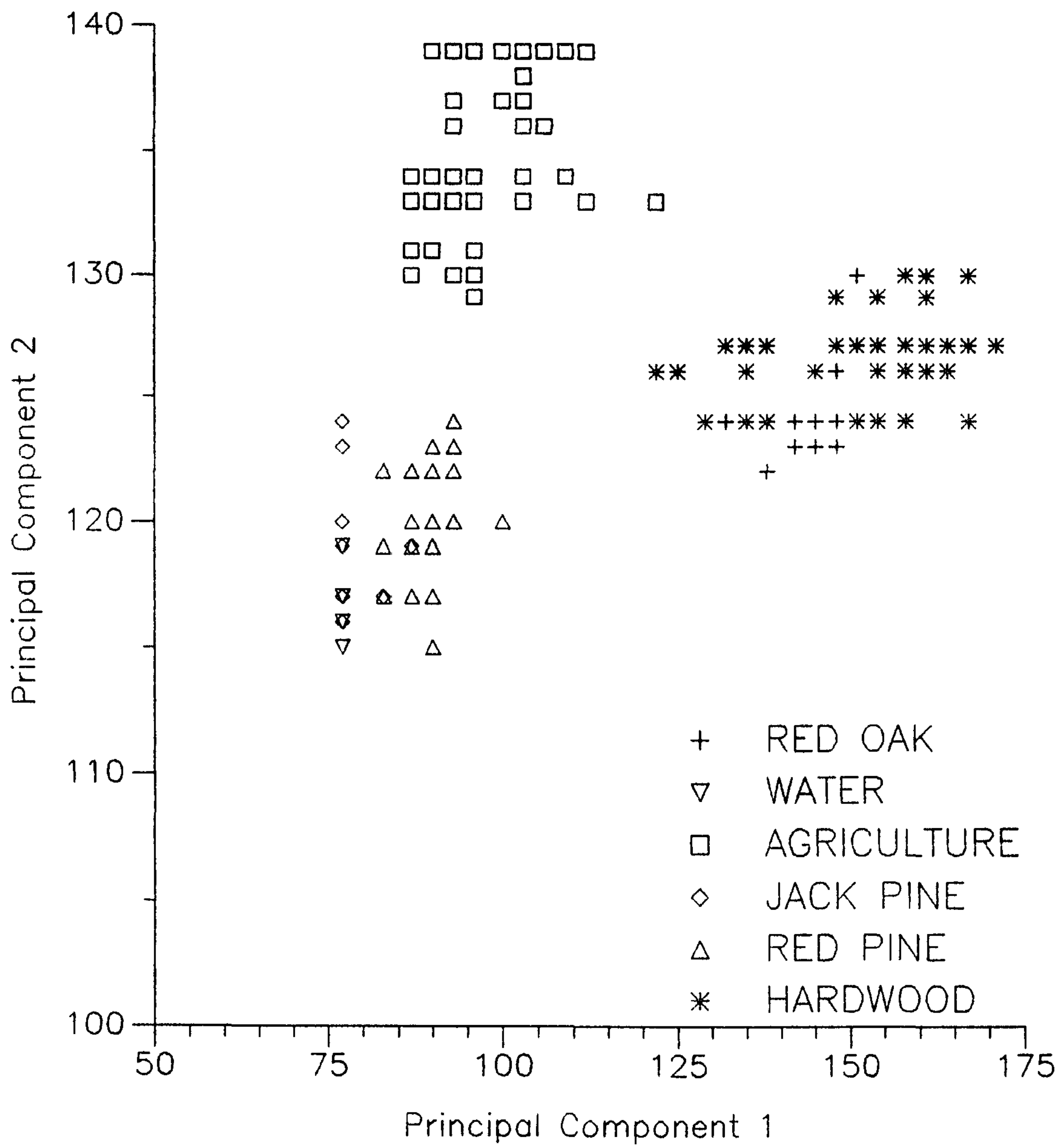


Figure 13 Feature Space for Principal Components 1 and 2

Figures 7 and 12 illustrate the marked correlation between the two visible-light bands 4 and 5 and the two near infrared bands 6 and 7. In Figure 12, the dispersion of the data is much higher than in Figure 7. Figure 7 shows that the agricultural areas have an elongated green-red reflectance cluster which overlays somewhat the hardwood 4 - 5 feature space. The elongated cluster of the agricultural areas can also be observed in Figures 8, 9, 10 and 11. On bands 6 and 7 (Figure 12) agricultural areas would be difficult to discriminate from the conifers, especially red pine.

On the principal component analysis of the imagery (Figure 13), agricultural areas would be discriminated very easily because of their high brightness values. Note the total lack of correlation between principal components 1 and 2. This is a characteristic of this type of statistical transformation.

In terms of the forest types of interest for the present project, forest and non-forest areas could be differentiated in the scene. Within the forest group, two general clusters could be separated, hardwood and conifers. This would warrant the division of the area into two strata. In bands 4 and 5 (Figure 7), this separation would be very difficult, if not impossible. Notice that when the bands in the infrared portion of the spectrum (bands 6 and 7) are included, the clusters became more evident. Between the two infrared bands, band 7 gives a better separation of the hardwood and conifer clusters than band 6. This can be seen by comparing Figures 8 and 9, where band 4 is plotted against bands 6 and 7, respectively and in Figures 10 and 11 where band 5 is plotted against bands 6 and 7, respectively.

In terms of subdividing the hardwood and conifer strata into hardwood verses red oak and red pine verses jack pine, some problems may occur. Obviously, the red oak pixels are always intermingled within the same cluster as the hardwood pixels. Although this spectral overlap is to be expected, on the printed number map it was possible to identify the pixels that composed the red oak stand. As a result of this spatial homogeneity, the samples could be selected and measured.

For the conifers, red pine and jack pine have some pixels that are mixed, but some of the jack pine pixels form a separate, albeit small, cluster. This can be detected mainly on the plots where the infrared bands (6 and 7) are present (Figures 8, 9, 10, 11 and 12). The best separation between these two conifer species would be obtained by using the principal component imagery, mainly PC 1 and PC 2, to run the classification. As seen on Figure 13, the two species form distinct clusters with the red pine having brighter values than the jack pine.

One aspect to be noticed on all of the plots shown is the cluster of water pixels. Water absorbs energy in the reflected infrared portion of the spectrum (Lillesand and Kiefer, 1987). Bands 6 and 7 are located in this region of the spectrum and, as a result, water should show zero or very low brightness values. As seen on Figures 8 through 12, there are some water pixels intermingled with the conifer cluster. In Figure 7 this is particularly noticiable. This is an example of pixels that were selected for training sites which were not "pure" water brightness values. The water bodies located within the study area are relatively small. As a result, in some of them, a portion of the selected pixels might have been located on the margins of the pond giving much higher brightness values. As seen in Figure 13, jack pine and water could not be separated if the principal components 1 and 2 were used for the classification. Even using the four bands, some confusion between jack pine and water would also be present. As occurred with the hardwood and red oak strata, in the printed number map, it was possible to identify stands of red pine and jack pine and to select specific samples.

The problem that occurs when overlap between two classes exists, as in the present case, is the occurrence of omission and commission errors. Omission errors occur when a pixel is not assigned to the appropriate class. Commission errors occur when a pixel is assigned to a class that it does not belong (Jensen, 1986).

Based only on the feature space plots shown, one can expect that the percentage of omission and commission errors might have been large on the classification performed on the study area. At this point a question arises: What would be the effect of these errors on the estimated total volume and its variance for the multiphase sampling technique?

The numeric results, in terms of number of sampling units (pixels), obtained from the computer classification are used on equations 16, 18 and 19. These equations are used respectively for the estimation of the variance of the estimated population mean for the i^{th} population stratum ($v(y_{i..})$); the estimation of the population mean (y_{st}) and its variance ($v(y_{st})$).

Just for the sake of a quick reference, the number of sampling units shown in Table 1 for the conifer and hardwood strata were changed by increasing and decreasing the original values by 5% and 10%. The values given by equation 16 were only slightly changed, but the estimated population mean and its variance were changed by roughly the same amount. In other words, if the numbers on Table 1 were increased or decreased 5% or 10%, the values given by equations 18 and 19 were also increased or decreased by approximately the same amount.

This was to be expected since the numbers on Table 1 are used to calculate the sampling fraction (w_i) for each stratum. The result of increasing or decreasing N_i, keeping N constant, can be visualized by examining equations 18 and 19.

It can be noticed from the discussion presented that, depending on the algorithm used for the classification, different results could be obtained for the estimated total volume and its variance. Of course, one cannot say if the calculated total volume and its variance is accurate or not since the population parameter values are not known. The choice of the classifier has to be done appropriately in order to reduce as much as possible the omission and commission errors. Previous experience with the classifier on the area of interest is important to consider when making the selection.

5.7 Time Measurements and Costs A comparison of the total time in hours necessary for the execution of each sampling procedures, Tables 8 and 21, shows a difference of only 1.6 hours. This can be considered as insignificant in relation to the total hours for both methods. A significant difference, however, exists between the total cost for each method. The reason for this difference is because in the multistage technique, the total hours were human hours and these were multiplied by the price per hour of $29.00. On the other hand, for the multiphase technique, 28.7% of the total hours was used for computer time which has a much lower price per hour. It must be emphasized that the computer wasn't only used for the classification of the imagery. It was also used to locate the selected sampling units (pixels) within the study area. For this study, one factor which greatly inflated the computer time was the selection of the data window that contained the study area from within the full scene on tape. The total number of computing hours required to classify the scene can be significantly reduced if the operator has experience in working with the equipment.

By analyzing Table 8 it is seen that, within each stratum, the field work had the highest cost as would be expected. For the conifer stratum, the total hours used for the selection of first stage sampling units is close to the hours used for field work. The reason for this is that the conifer stratum was the first one measured in the process. As more experience was gained, the total hours used for the selection of first stage sampling units decreased.

The time used for the selection of the second stage sampling units also varied among the strata. This variation is associated with the place were the primary unit was selected. The primary units selected for the hardwood stratum were located on relatively homogeneous portions of the stratum. As a result, the time spent on measuring the area of the stratum contained within each secondary unit was low. This was not the case for the conifer and red oak strata. The measurement of the area of each stratum within the secondary units had to be done more carefully since parts of other strata (hardwood, agricultural areas) were also contained within the secondary units.

The selection of the third stage units also used a relatively large number of hours. Most of the time was used in the evaluation of the percent crown closure within the 121 tertiary sampling units, within each selected secondary unit.

With the field measurements, the time used for measuring plots in the conifer stratum was greater than that for travel between plots. This is because most of the plots within the conifer stratum were concentrated in one area. As a result, the time for moving between plots was reduced. For the hardwood stratum, the difference between measurement time and travel was small. The selected plots within the hardwood stratum were located on an area of more difficult access and also were not concentrated in only one place, as occurred with the conifer stratum. For the red oak stratum, there was a large difference between the time used for measuring the plots and for travel time. The red oak stratum is concentrated in only one portion of the study area, and the selected plots were relatively close. The reason for the larger difference in time was the fact that the stand had been thinned recently and it was extremely difficult to walk due to the slash left on the ground.

As related to the multiphase technique (Table 21), it can be seen that the time used for sampling selection on the hardwood and conifer strata is relatively high. As already stated, the selection of the samples was done on the number map. The first step in this process was to identify the clusters of pixels that were classified within each stratum of interest. The coordinates of the clusters were identified and the selection process was started. During the selection process, a large number of selected sampling units had to be discarded because they did not fall within the forest type of interest.

The hours required for field measurements in the multiphase method were much less, since fewer samples were measured. Table 21 shows that, for the conifer stratum, the total time required for measuring the plots on the red pine and jack pine plantations (16.1 hours) was practically the same as for the multistage method (16.7 hours). The difference occurs in the time used for measuring the plots and to move between plots. For the multistage technique twelve plots were measured within the conifer stratum, and ten for the multiphase technique (five within the red pine stratum and five within the jack pine stratum). This difference accounts, in part, for the larger time needed to measure plots for the multistage technique. In terms of moving between plots, the multiphase technique used more time, even though fewer plots were measured. The reason is that the multistage technique tends to concentrate the plots in one region. In the multiphase technique, the plots are randomly selected within each stratum and there is a tendency for them to be spread out. As a result, the time used to move between plots tends to be larger than for the multistage technique.

The difference between the time used for measuring the plots and to move between plots is quite large for the hardwood stratum. Besides the distance between the plots, some of them were located in areas of difficult access which increased the time to reach the plot.

For the red oak stratum, the time used in measuring the plots and in moving between plots were practically the same. Of the six plots measured within this stratum by the multi-phase technique, two were located in an area that was thinned. The other plots were located in areas of easy access and were not thinned. This contributed to the reduction of both the time for measuring and the time for moving between plots.

The average time required for measuring the 36 plots in the multistage technique was of one hour and fifty four minutes. For the multiphase technique, the average time for measuring the 21 plots was of one hour and fourty eight minutes. The difference of only six minutes can not be considered as significant, as a result, for the conditions of the present work, both methods were equivalent in terms of the time spent for field work.

5.8 **Number of Samples** The equations presented in Appendix A are used to determine the number of samples to be measured for each method. For the multistage technique, the determination of the required number of samples to be measured at each stage is based on optimum allocation (Langley, 1975). Optimum allocation can be achieved by minimizing variance for a specific cost or by minimizing cost for a specific variance.

For the multiphase technique it is necessary to first calculate the sampling fractions u and v for the first and the second phases, respectively. The calculation of only two sampling fractions instead of one for each of the i^{th} phase one strata and four for the second phase, implies proportional allocation of the sampling units in each stratum. This treatment is given by Johnston et al, 1983.

Frayer (1979) discussed three possible ways of determining the sample size n and explicit sampling fractions for the first and second phases, based on the optimality criterion as given above. The first approach assumes that the sampling fractions are equal and unspecified. The second assumes that the sampling fractions are unequal and specified, and the third approach assumes that the sampling fractions are unequal and unspecified. This last case requires the use of non-linear programing techniques to solve the problem.

In order to estimate the costs associated with the calculations of the sampling fractions u and v, one can subdivide Table 21 into three sections. The first section would refer to the costs associated with computer classification. This can be considered as the cost C1 shown on the equation to calculate the sampling fraction u. The second section would refer to the costs associated with the calculation of both u and v. These may be referred as the costs C2 on the equations on Appendix A, and include the costs related to the conifer and the hardwood strata, presented in Table 21. The third section would refer to the costs associated with the calculation of v. As seen in Appendix A, these costs would be associated with the term C3 and would refer to the field and area measurements costs, for each of the four strata.

The calculations of the optimum number of samples on each stage of the multistage technique, and of the sampling fractions for the multiphase technique, would not be possible on the present case. This is because the work was performed by one person on a small scale. As the estimated variances are large and the costs are rather small, the number of samples to be measured for each method, calculated by applying the equations on Appendix A, are not realistic. The number of sample plots measured for each technique was arbitrarily established, based on the time and funds available for the field work.

CHAPTER 6

SUMMARY AND CONCLUSIONS

This study's main objective was to evaluate two multilevel sampling procedures for forest inventory in northern Michigan. A multistage and a multiphase sampling technique were evaluated.

Multistage sampling uses unequal probability sampling, also referred to as sampling with probability proportional to size (PPS). This method requires analog imagery for the area of interest. Multiphase sampling requires digital imagery for the area of interest; it is based on stratified random sampling and uses two phase sample rather than simple random sampling in each stratum (Johnston, 1982).

The following conclusions should be emphasized:

a) The multistage technique produced estimates of the population total that were 5 times larger than the multiphase technique. This discrepancy is basically due to the high variability in the volume found in the population, which resulted in different estimates of total volume for some of the strata (e.g. hardwood). The differences in structure of both methods also contributed to differences in estimated population totals.

b) Although the results must be viewed with caution due to the large variances encountered and the small sample sizes, the multiphase technique was more precise in estimating the population total and less costly than the multistage technique. The multiphase technique gave more precise estimates of the total volume than the multistage technique for all individual stratum, except for the hardwood stratum.

c) Contrary to expectations, the correlation between the the size variable and the predicted volume for the second stage of the multistage technique was negative. This is an indication that, in the present application of the method, estimated total area occupied by forest type within the secondary sampling unit is not an appropriate size variable to calculate the selection probabilities.

d) The size variable used for the calculation of the selection probabilities for the third stage (percent crown closure) was not appropriate either. The correlations between percent crown closure and estimated volumes for the third stage were positive for all three strata, as expected. But the values of such correlations were low and not highly significant. A high percent crown closure is not necessarily associated with high plot volume.

e) Although the correlations between the size variable and the predicted volume for the second stage were negative for all strata, the second stage gave the smallest contribution to the variance of the estimated total at all strata.

f) The changing of the selection probabilities for each stage of the multistage technique influenced the estimated total volume and its variance. This occurred because the probabilities are an integral part of the equations used to estimate those parameters. No rule can be established for the effect of changing the selection probabilities on the estimated total volume and its variance, however. Care should be taken when measuring the size variables, specially for the second and third stages, since changes in the probabilities at these stages reflect more on the estimated variance of the total than the probabilities of the first stage.

g) The size of the primary sampling unit (PSU) used in the present study was not convenient for the following reasons: it produced only two PSU's for the red oak stratum; it resulted in secondary sampling units having zero values for the size variable; and for the conifer stratum, although all PSU's contained some conifer plantations, it produced several secondary sampling units with zero values for the size variables. These zero readings in turn caused great variability in the strata's selection probabilities. This variation, in turn, caused the negative correlations between the size variable and the predicted volume for the second stage. There is no rule of thumb to determine the size of the primary units. In each application of the method, one has to evaluate the most appropriate size as a function of the number of stages being utilized, the scale of the imagery at each stage, and the size of the field plots.

k) The use of the principal component imagery to classify the study area scene would have been better since both species of conifer could be better discriminated, and because agricultural areas and hardwoods formed separate clusters. Although the discrimination between red oak and mixed hardwood was very difficult, not only on the principal component imagery but also on all the other band combinations, the red oak stand could be identified on the number map of the classified scene. The discrimination between jack pine and water would not be possible on the principal component imagery.

The selection of a sampling procedure for forest inventory depends, among other things, on familiarity and experience with the method. Multilevel sampling methods are very useful when large areas are to be inventoried, and the use of remotely sensed imagery in the form of aerial photographs and, more recently, satellite imagery is necessary. The methods evaluated by this research take advantage of the use of remotely sensed imagery in either its analog or digital form.

The imagery used for both methods was simply what was available at the start of the project. No provisions for taking large-scale aerial photographs were available. Such photographs are very helpful for multilevel sampling.

The scales of the imagery used for the study had an influence on the determination of the size of the sampling units at each stage of the multistage method. In situations where the number of levels of imagery is fixed, such as in the present application, the size of the primary sampling units has an influence on the size of the field plots (last stage). If the size of the PSU's is increased, the size of the field plots will also increase and vice-versa. A small field plot is convenient to measure but, depending on the scale of the last level of imagery, measurement of the size variable may be difficult. This evaluation - size of PSU vs size of field plots - has to be done carefully in situations were stratification of the area to be inventoried is necessary. The smallest stratum in the area should be considered first and the best judgement, relative to the size of the primary sampling units, should be taken.

Besides the size of the strata, another variable to be considered in defining the size of the sampling units at each stage is the variation in the volume. If this information is available, or can be anticipated in some way, enough primary sampling units should be placed within each stratum to ensure the selection of the required number in order to attain the specified precision of the estimate. If, on the other hand, such information is not available, the size of the primary sampling units should be such to allow for more than two PSU's to be completely contained within the strata of smaller areas.

The correlation between the size variable and the estimated volume for the second and third stages was analyzed in some detail. Although the correlations between those two variables for the second stage were negative for all three strata, this does not mean that this will always happen. It is probable that in situations where the secondary sampling units are larger, a positive correlation will be established.

Another aspect to consider is the fact that the size variable "area" is inferior to an estimate of volume, which would be the ideal size variable to use. "Area" however is superior to ocular volume estimates by inexperienced assessors (Loetsch et al, 1963). An alternative to percent crown closure for use if multistage sampling is chosen could be crown area. This is obtained by multiplying the mean percent crown cover by the stratum area (Loetsch et al, 1963). When the scale of the imagery permits, the number of emergent canopies could also be tried. If an area is going to be reinventoried and informations on the volume per unit area is available from previous inventories, these informations can be used as size variables. The advantage of using past informations is their highly correlation with the variable of interest (new estimation of volume).

The multistage sampling technique evaluated has a rather complex structure and, as seen from the previous comments, the definition of the size of the PSU has an important effect on the results obtained mainly in situations where stratification of the area is done. Some aspects for future research could be pointed out related to the multistage technique: perform simulated studies in conditions where stratification by forest types is recommended to define the most appropriate size for the PSU's, when the number of levels of imagery is fixed. The same study could be done but, without the limitations on the number of levels of imagery in order to define both, the size of the PSU and the most economic number of stages to be considered. For a not so complex study, the last condition could be evaluated in a situation when the forest is homogeneous or no stratification is required. The evaluation of alternative size variables (not predicted volume) may also be considered as another line of study.

The application of the multistage method in situations where stratification of the area is recommended may impose some limitations to the use of the method. Consider for instance the conifer stratum on the present study. This stratum was formed of small plantations of red and jack pine spread all over the study area and of large plantations of red pine concentrated in the south east portion of the area. Due to this characteristic of this stratum, it was considered as one and not subdivided into red and jack pine. The field measurements were made on the large plantations of red pine since the PSU contained within these plantations was selected twice, due to its large selection probability. As a result, the estimated total volume of the conifer stratum is reflecting mostly the availability of red pine. The multiphase technique, on the other hand, probably gave a more realistic appraisal of total volume on the conifer stratum because the structure of the method allowed for measurements to be made on both red and jack pine plantations.

The multiphase sampling technique evaluated has a much simpler structure since it is based on stratified random sampling. As the method requires the use of the digital representation of the image, a limitation for the use of this method is, of course, the lack of facilities to digitally process the image. If such facilities are available, one may ask the question: "What would be the most appropriate method for the classification of the imagery?" The sensitivity analysis performed on the data obtained from the digital processing of the imagery on the present study showed that the effect of changing the number of pixels classified in each stratum of interest, has much less effect on the estimate of the population parameter than did the changing of the selection probabilities of the multistage method. This means that either supervised or unsupervised classification algorithms could be used to classify the image. Of course, the algorithm that gives the least percentages of omission and comission errors for a particular case should be the one chosen. In order to have this information available before the application of the multiphase sampling technique, it is necessary to know which classification algorithm would give a better accuracy for a certain situation. This could be an area of future research. Taking the several cover types that occur in Michigan and performing both supervised and unsupervised classifications, using imagery from different seasons, to evaluate which algorithms would be more accurate. If these data are available in advance, the future users of the multiphase sampling technique could benefit from them.

Based solely on the conditions established for the application of both sampling techniques, it is clear from the results that the multiphase method was more adequate for the inventory of the particular area selected. This does not mean that the other method should not be considered for inventories in northern Michigan. If a consulting forester or an organization wants to use a multilevel sampling technique for an inventory it is not advisable to jump right into the multiphase technique evaluated, just because it was more precise and less costly. A series of variables and analysis, as previously discussed, have to be considered before making the final selection.

APPENDIX A

EQUATIONS FOR THE OPTIMUM ALLOCATION OF SAMPLES AT EACH STAGE AND FOR CALCULATION OF THE SAMPLING FRACTIONS

A. Optimum Allocation at Each Stage

A.1 First Stage

$$m = (D^*)\sqrt{(E_1/D_1)}/(\sqrt{E_1 D_1} + \sqrt{E_2 D_2} + \sqrt{E_3 D_3})$$

A.2 Second stage

$$n = \sqrt{E_2 D_1}/\sqrt{E_1 D_2}$$

A.3 Third stage

$$t = \sqrt{E_3 D_2}/\sqrt{E_2 D_3}$$

$$E_1 = \sum_{i=1}^{M} P_i (V_i/P_i - V)^2$$

$$E_2 = \left(\sum_{j=1}^{N_i} P_{ij} (V_{ij}/P_{ij} - V_i)^2\right)/P_i$$

$$E_3 = \left(\sum_{i=1}^{M} \sum_{j=1}^{N_i} \sigma_{3ij}/\sqrt{N_i}\right) \sum_{j=1}^{N_i} \sigma_{3ij}\sqrt{N_i}$$

D_1 = average cost of measuring a first-stage unit,

includes cost of ennumerating the predictions.

D_2 = average cost of measuring a second stage unit, includes cost of ennumerating the predictions.

D_3 = average cost of measuring a third stage unit, includes travel costs.

D^* = expected cost of the survey

B. Calculation of Sampling Fractions

B.1 Phase one Sampling Fraction

$$u = \sqrt{C_1 S^*_2} / \sqrt{C_2 S^*_1}$$

B.2 Phase two Sampling Fraction

$$v = \sqrt{C_2 S^*_3} / \sqrt{C_3 S^*_2}$$

B.3 Number of Samples

$$n = C^* / (C_1 + C_2 u + C_3 uv)$$

$$S^*_1 = s^2 - s' \qquad S^*_2 = s' - S^*_3$$

$$S^*_3 = \sum_{i=1}^{L} \sum_{j=1}^{I_i} (m_{ij} n_i / m_i n) / s_{ij}^2$$

$$s' = \sum_{i=1}^{L} (n_i / n) s_i^2$$

$$s^2 = \frac{Nn}{N-1} \left(v(y) - \sum_{i=1}^{L} (n_i / n)^2 v_i \right)$$

$$v(y) = \frac{n(N-1)}{N(n-1)} \left[\frac{N-n}{n(N-1)} \sum_{i=1}^{L} (n_i/n)(y_i - y)^2 + \right.$$

$$\left. \frac{N-n}{n^2(N-1)} \sum_{i=1}^{L} (n_i - 1)s_i^2 + \sum_{i=1}^{L} (n_i/n)\left((n_i/n) - \frac{(N-n)}{n(N-1)}\right) v_i \right]$$

$$v_i = \frac{n_i - 1}{n_i} \sum_{j=1}^{I_i} \left(\frac{m_{ij}-1}{m_i - 1} - \frac{b_{ij}-1}{n_i - 1} \right) \frac{m_{ij}}{m_i b_{ij}} s_{ij}^2 +$$

$$\frac{n_i - m_i}{n_i(m_i - 1)} \sum_{j=1}^{I_i} (m_{ij}/m_i)(y_{ij} - y_i)^2$$

$$s_i^2 = \frac{m_i n_i}{(n_i - m_i)} \left[v_i - \frac{1}{m_i} \sum_{j=1}^{I_i} (m_{ij}/m_i) s_{ij}^2 ((m_{ij}/b_{ij}) - 1) \right]$$

$$s_{ij}^2 = 1/(b_{ij}-1) \sum_{k=1}^{b_{ij}} (y_{ijk} - y_{ij})^2$$

C^* = expected cost of the survey

C_1 = fixed cost per phase one unit

C_2 = cost per second phase unit

c_3 = cost per third phase unit

s^2 = estimated population variance

s_i^2 = estimated population variance of the i^{th} phase one stratum

s_{ij}^2 = estimated population variance of the $(ij)^{th}$ phase two stratum

The other terms are as defined on the text.

APPENDIX B

Selection Probabilities for the Hardwood, Conifer and Red Oak Strata

A. Hardwood Stratum

A.1 Original Probabilities:

P_i	P_{ij}	P_{ijk}
.09987	.06253	.00809
.07918	.08348	.01059
	.11689	.01059
	.14945	.00873
		.00989
		.00989
		.01061
		.01061
		.00687
		.00898
		.00804
		.00898

A.2 Changed Probabilities:

P_i

Trial#1	Trial#2	Trial#3	Trial#4	Trial#5
.10531	.10449	.10524	.10953	.09192
.0835	.06779	.08345	.08685	.07288

P_{ij}

Trial#1	Trial#2	Trial#3	Trial#4	Trial#5
.06867	.07354	.05863	.07051	.06898
.09168	.08033	.09568	.07702	.0921
.13817	.15971	.1353	.13855	.15727
.13209	.12493	.10583	.10837	.12301

P_{ijk}

Trial#1	Trial#2	Trial#3	Trial#4	Trial#5
.00925	.00945	.00954	.00928	.00706
.00925	.00945	.01209	.01176	.00963
.01171	.00945	.00954	.01176	.00963
.00763	.0099	.00995	.00757	.0074
.00881	.00874	.00878	.00874	.00854
.01116	.00874	.00878	.01107	.00854
.01169	.01196	.00948	.01179	.00939
.00923	.01196	.00948	.01179	.00939
.00554	.00819	.00569	.00558	.00563
.00831	.00828	.00822	.00826	.00831
.00733	.00731	.00726	.00923	.00733
.00928	.00828	.00919	.00826	.00929

B. Conifer Stratum

B.1 Original Probabilities:

P_i	P_{ij}	P_{ijk}
.10280	.10640	.01027
.10280	.18470	.01941
	.08790	.01941
	.10640	.01142
		.01007
		.01007
		.01931
		.01158
		.01931
		.01941
		.01256
		.00571

B.2 Changed Probabilities:

P_i				
Trial#1	Trial#2	Trial#3	Traial#4	Trial#5
.11785	.09432	.11143	.09521	.11276
.11785	.09432	.11143	.09521	.11276

P_{ij}

Trial#1	Trial#2	Trial#3	Trial#4	Trial#5
.09854	.11303	.09718	.11341	.11727
.17108	.19623	.16872	.16110	.20359
.09953	.07642	.09816	.09372	.09691
.09854	.11303	.09718	.11341	.11727

P_{ijk}

Trial#1	Trial#2	Trial#3	Trial#4	Trial#5
.00806	.01259	.00787	.00816	.01209
.01728	.02174	.02135	.01748	.01648
.02189	.02174	.01685	.02214	.02088
.00999	.01249	.01010	.01276	.01273
.00866	.01118	.01145	.01142	.00871
.01133	.00855	.00875	.01142	.00871
.02149	.01722	.01722	.01643	.01704
.00885	.01457	.01457	.00885	.01442
.02149	.01722	.02252	.02149	.02228
.01728	.01716	.02135	.01748	.02088
.01037	.01487	.01011	.01049	.00989
.00806	.00343	.00787	.00350	.00769

C. Red Oak Stratum

C.1 Original Probabilities:

P_i	P_{ij}	P_{ijk}
.84395	.16319	.00793
.84395	.23511	.00887
	.23511	.00887
	.16319	.00741
		.00741
		.00840
		.00840
		.00840
		.00840
		.00887
		.00793
		.00887

C.2 Changed Probabilities

P_i

Trial#1	Trial#2	Trial#3	Trial#4
.86859	.84395	.84395	.81566
.86859	.84395	.84395	.81566

P_{ij}

Trial#1	Trial#2	Trial#3	Trial#4	Trial#5
.17365	.14681	.17168	.15421	.14682
.25019	.25852	.24736	.27156	.21154
.25019	.25852	.24736	.27156	.21154
.17365	.14681	.17168	.15421	.14682

P_{ijk}

Trial#1	Trial#2	Trial#3	Trial#4	Trial#5
.00721	.00900	.00902	.00724	.00911
.00913	.00900	.00807	.00821	.00815
.00817	.00805	.00902	.00821	.00815
.00648	.00840	.00635	.00839	.00838
.00648	.00643	.00830	.00839	.00838
.00747	.00741	.00928	.00937	.00739
.00747	.00939	.00733	.00937	.00739
.00947	.00939	.00733	.00740	.00739
.00947	.00939	.00928	.00937	.00739
.00913	.00805	.00902	.00821	.00815
.00913	.00900	.00712	.00917	.00719
.00817	.00900	.00807	.00821	.00911

APPENDIX C

List of Scientific Names of Trees

Common Name	Scientific Name
American Beech	*Fagus grandiflora Ehrh.*
Basswood	*Tilia americana L.*
Bigtooth Aspen	*Populus grandidentata Michs.*
Blach Cherry	*Prunus serotina Ehrh.*
Elm	*Ulmus spp.*
Iron wood	*Ostrya virginiana (Mill.) K.Koch*
Jack Pine	*Pinus banksiana Lamb.*
Oak	*Quercus spp.*
Red Maple	*Acer rubrum L.*
Red Oak	*Quercus rubra L.*
Red Pine	*Pinus resinosa Ait.*
Sugar Maple	*Acer saccharum Marsh.*
Trembling Aspen	*Populus tremuloides Michx.*
White Ash	*Fraxinus americana L.*

APPENDIX D

TALLY SHEET

FIELD COLLECTION DATA SHEET
Summer 1987

Date: / /87 Plot#:_______ BAF:______

Location:_______________________________________

Time start: : Finish: :

Point# 1				Point# 2			
Tree#	Spp	DBH	MerchHT	Tree#	Spp	DBH	MerchHt
1				1			
2				2			
3				3			
4				4			
5				5			
6				6			
7				7			
8				8			
9				9			
10				10			
11				11			
12				12			
13				13			
14				14			
15				15			
16				16			
17				17			
18				18			
19				19			
20				20			

Comments:_______________________________________

APPENDIX E

A. Volume per Plot in cubic feet/acre for the Multistage Sampling Technique

A.1 Hardwood Stratum

Plot#	Volume
1.1.1	2,047.05
1.1.2	2,206.81
1.1.3	2,066.79
1.2.1	2,589.14
1.2.2	3,455.34
1.2.3	3,100.82
2.1.1	1,537.66
2.1.2	1,431.21
2.1.3	1,772.23
2.2.1	2,216.26
2.2.2	2,176.66
2.2.3	1,821.06

A.2 Conifer Stratum

Plot#	Volume
1.1.1	3,042.66
1.1.2	3,668.98
1.1.3	2,185.42
1.2.1	2,963.12
1.2.2	2,137.78
1.2.3	2,532.41
2.1.1	1,976.81
2.1.2	1,924.06
2.1.3	3,123.94
2.2.1	2,568.85
2.2.2	1,769.06
2.2.3	1,414.91

A.3 Red Oak Stratum

Plot#	Volume
1.1.1	1,833.90
1.1.2	2,654.62
1.1.3	2,659.69
1.2.1	2,171.66
1.2.2	2,419.89
1.2.3	2,123.22
2.1.1	2,033.44
2.1.2	1,764.94
2.1.3	2,166.95
2.2.1	3,340.35
2.2.2	2,850.24
2.2.3	2,533.65

B. Volume Per Plot in cubic feet/acre for the Multiphase Sampling Technique

B.1 Hardwood Stratum

Plot#	Hardwood	Red Oak
1	2,423.86	1,385.75
2	1,615.05	2,506.85
3	1,243.82	2,301.51
4	3,204.58	1,958.07
5	1,298.34	1,989.80
6		2,937.91

B.2 Conifer Stratum

Jack Pine	Red Pine
1,945.29	3,517.55
1,245.04	3,109.69
1,598.59	1,995.50
1,421.78	1,656.17
1,325.95	2,847.39

LITERATURE CITED

Albert, D.A.; Denton, S.R.; Barnes, B.V. 1986. Regional landscape ecosystem of Michigan. School of Natural Resources, The University of Michigan, 32 p.

Aldred, A.H. 1980. Forest inventory by multistage remote sensing. In: Remote Sensing Symposium, Proceedings, Great Lake Forest Research Center, Sault Ste. Marie, Ontario, Canada, pp. 77 - 84.

Anderson, J. E. 1979. Multistage variable probability forest volume inventory. NASA/ERL Report #179, 32 p.

Avery, G.; Meyer, M.T. 1959. Volume tables for aerial timber estimating in northern Minnesota. Lake States Forest Experimental Station, station paper #78, 21 p.

Beers, T.W. 1964. Composite hardwood volume tables. Purdue University, Agricultural Experiment Station, Research Bulletin 787, 12 p.

Beers, T.W.; Miller, C.I. 1966. Horizontal point sampling tables. Purdue University, Agricultural Experimental Station, Research Bulletin 808, 80 p.

Bichford, C.A. 1961. Stratification for timber cruising. In: Journal of Forestry, 59 (10), 761 - 763.

Bichford, C.A.; Mayer, C.E. and Ware, K.D. 1963. An effective sampling design for forest inventory: The northeastern forest resurvey. In: Journal of Forestry, 61 (11):826 - 833.

Buchanan, D.E. 1985. Soil survey of Lake and Wexford Counties Michigan. USDA Soil Conservation Service, 135 p.

Chapelle, D.E. 1985. The research process in natural resources. Lecture notes for the course RD/FOR 855, Michigan State University, Forestry Department, 156 p.

Cochran, W.G. 1977. Sampling Techniques. John Willey & Sons, Third Edition, 428 p.

De Vries, P.G. 1986. Sampling theory for forest inventory - a teach yourself course. Spring-Verlag, 399 p.

Donker, N.H.W.; Mulder, N.J. 1977. Analysis of MSS digital imagery with the aid of principal component transformation. In: ITC Journal, Vol. 3, pp. 434 - 467.

Frayer, W.E. (Editor) 1979. Multilevel sampling designs for resource inventories. Report to US Forest Service, Rocky Mountain and Range Experiment Station, Fort Collings, CO, 113 p.

Frayer, W.E. 1981. Multi-level sampling designs for resource inventories. In: XVII IUFRO World Congress, Proceedings, Division 9, Japan, pp. 169 - 173.

Freese, F. 1962. Elementary forest sampling. Agriculture Handbook 232, US Department of Agriculture, 91 p.

Gialdini, M.; Titus, S.; Nichols, J.; Thomas, R. 1975. The integration of manual and automatic image analysis technique with supporting ground data in a multi-stage sampling frame work for timber resource inventories: three examples. In: Proceedings of the NASA Earth Resources Survey Symposium, Houston, Texas, Lyndon B. Johnson Space Center, pp. 1377 - 1387.

Goodenough, D.; Shlien, S. 1974. Automatic classification methodology. Canada Center for Remote Sensing, Department of Energy, Mines and Resources, Research Report 74-1, 8 p.

Hall, J.A.; Hales, M.T. 1979. Multistage inventory technique case study - McCoin Valley. In: Seventh Biennial Workshop on Color Aerial Photography in the Plant Sciences and Related Fields, Proceedings, Davis, California, pp. 119 - 124.

Harding, R.A.; Scott, R.B. 1978. Forest inventory with LANDSAT phase II - Washington forest productivity study. State of Washington Department of Natural Resources, Olympia, Washington, 221 p.

Harris, J.W.E.; Dawson, A.F. and Brown, R.G. 1983. Evaluation of mountain pine beetle damage using aerial photography taken with a hand-held 70 mm camera, Gold Bridge - Clinton, B.C., 1981. Canadian Forest Service, Pacific Forest Research Center, BC-X-245, 15 p.

Hegyi, F. 1980. A new approach to forest inventory using remote sensing. In: Remote Sensing Symposium, Proceedings, Great Lake Forest Research Center Sault Ste. Marie, Ontario, Canada, pp. 94 - 99.

Heller, R.C.; Wear, J.F. 1969. Sampling forest insect epidemics with color films. In: Sixth International Symposium on Remote Sensing of Environment, Proceedings, Vol. II, pp. 1157 - 1167.

Hesse, R.A. 1987. Private communication.

Hudson, D.W. 1986. Classification of coniferous forest cover types using LANDSAT MSS digital data. Ph.D. Dissertation, Department of Forestry, Michigan State University, East Lansing, 251 p.

Hush, B. 1971. Planning a forest inventory. FAO Forestry and Forest Products Studies, #17, 120 p.

Hush, B.; Miller, C.I. and Beers, T.W. 1982. Forest mensuration. John Willey & Sons, Third Edition, 402 p.

Hutchinson, I.D. 1978. An example of the use of two phase sampling design for reconnaissance inventory in tropical forest. In: New Zealand Journal of Forestry, 23(1):95 - 106.

Jensen, R.J. 1986. Introductory digital image processing - a remote sensing perspective. Prentice-Hall, 379 p.

Jeyaratnam, S.; Bowden, D.C.; Graybill, F.A. and Frayer, W.E. 1984. Estimation in multiphase designs for stratification. Forest Science, 30(2):484 - 491.

Johnston, D.C. 1982. Theory and application of selected multilevel sampling designs. Ph.D. Dissertation, Department of Forest and Wood Sciences, Colorado State University, Fort Collins, 197 p.

Johnston, D.C.; Frayer, W.E. 1983. Stratified two phase sampling as a base for monitoring multiresources. In: Renewable Resource Inventories for Monitoring Changes and Trends, Proceedings of an International Conference. Edited by John F. Bell and Toby Atterbury. Corvallis, Oregon, August, 15-19, pp. 448 - 451.

Johnston, D.C. 1987. Private conversation.

LaBau, V.J.; Schreuder, H.T. 1983. A multiphase, multiresource inventory procedure for assessing renewable natural resources and monitoring change. In: Renewable Resource Inventories for Monitoring Changes and Trends, Proceedings of an International Conference.

Edited by John F. Bell and Toby Atterbury. Corvallis, Oregon, August 15-19, pp. 456 - 459.

Langley, P.G. 1969. New multistage technique using space and aircraft imagery for forest inventory. Symposium on Remote Sensing of Environment, Proceedings, 6(2):1179-1183.

Langley, P.G. 1971. The benefits of multistage variable probability sampling using space aircraft imagery. In: Application of Remote Sensing in Forestry, joint report by IUFRO working group "Application of Remote Sensors in Forestry", IUFRO Section 25, Germany.

Langley, P.G. 1975. Multistage variable probability sampling: theory and use in estimating timber resources from space and aircraft photography. Ph.D. Dissertation, Wildland Resource Science, University of California, Berkeley, 101 p.

Langley, P.G. 1978. Remote sensing in multistage, multiresource inventories. In: Integrated Inventories of Renewable Natural Resources, Proceedings of the Workshop, Rocky Mountain Forest and Range Experiment Station, general technical report RM-55, pp. 205 - 208.

Lee, D.C.L.; Hernandes, P.F.; Shimabukuro, Y.E.; Assis, O.R. de; Medeiros, J.S. de 1984. Forest inventory using multistage sampling with probability proportional to size. Instituto de Pesquisas Espaciais - INPE, publicacao #3084-PRE/494, Sao Paulo.

Li, H.G.; Schreuder, H.T.; Bowden, D.C. 1984. Four phase sampling estimation for the Alaska survey. In: Inventorying Forest and other Vegetation of the High Latitude and High Altitude Regions. Proceedings of an International Symposium, Society of American Foresters Regional Technical Conference, July 23-26, Fairbanks, Alaska, pp. 61 - 67.

Lillesand, T.M.; Kiefer, R.W. 1987. Remote sensing and image interpretation. John Willey & Sons, 721 p.

Loetsch, F.; Haller, K.E. 1964. Forest inventory. BLV Verlagsgesellschaft Munchen Basel Wien, Vol. 1, 436 p.

Lusch, D.P. 1987. Private conversation.

Lusch, D.P. 1988. Private conversation

MacClean, C.D. 1972. Improving inventory volume estimate by double sampling on aerial photographs. In: Journal of Forestry, 70(2):748 - 749.

Mattila, E. 1984. Survey of forest resources of Finish Lapland using multiphase systematic sampling. In: Inventorying Forest and other Vegetation of the High Latitude High Altitude Regions. Proceedings of an International Symposium. Society of American Foresters, Regional Technical Conference, 23-26 July, Fairbanks, Alaska, pp. 55 - 60.

Myers, W.L.; Shelton, R.L. 1980. Surveys methods for ecosystem management. John Willey & Sons, 403 p.

Nichols, J.D.; Gialdini, M. and Jaakkola, S. 1973. A timber inventory based upon manual and automated analysis of ERTS-1 and supporting aircraft data using multistage probability sampling. In: Third Earth Resources Technology Satellite 1, Symposium, Section A, NASA SP-351, Goddard Space Flight Center, Washington DC, pp. 145 - 157.

Nichols, J.D.; Harding, R.A.; Scott, R.B.; Edwards, J.R. 1976. Forest inventory of western Washington by satellite multistage sampling. In: Proceedings, American Society of Photogrammetry, Fall Convention, pp. 180 - 217.

Norton, J. 1987. Private conversation.

Nyyssonen, A. 1976. Objectivos y planeamiento de inventarios forestales. Curso FAO/FINLANDIA de entrenamiento en inventario forestal, Finlandia, pp. 64 - 80.

Peterson, D.L.; Noren, D.; Gnauck, G. 1983. Methods and results of three unequal probability multistage sampling designs for timber volume in the pacific northwest. In: Renewable Resource Inventories for Monitoring Changes and Trends. An International Conference, August 15-19, Corvallis, Oregon pp. 326 - 329.

Shiue, C.J.; John, H.H. 1962. A proposed sampling design for extensive forest inventory: double systematic sampling for regression with multiple starts. In: Journal of Forestry 60(9):607 - 610.

Syan-Wittgenstein, L. 1961. Phenological aids to species identification on airphotographs. Forest Research Branch, Technical Note #104, Canada Department of Forestry, 26 p.

Temu, A.B.; Philip, M.S. 1981. Sampling woodland for fuelwood. In: Proceedings of Forest Resource Inventory, Growth Models, Management Planning and Remote Sensing. XVII World Congress, IUFRO, Ed. Masahisa Nishizawa, Kyoto, Japan, pp. 272 - 279.

Titus, S.; Gialdini, M.; Nichols, J. 1975. A total timber resource inventory based upon manual and automated analysis of LANDSAT 1 and supporting aircraft data using stratified multistage sampling techniques. In: Tenth International Symposium on Remote Sensing of Environment, Proceedings, pp. 1093 - 1099.

Wear, J.F.; Pope, R.B. and Lauterbach, P.G. 1964. Estimating beetle-killed Douglas-fir by aerial photo and field plots. In: Journal of Forestry, 62(5):309 - 315.

Wert, S.L. 1968. Douglas-fir beetle survey with color photos. In: Photogrammetric Engineering, 34(12):1234 - 1248.

Wert, S.L. 1969. A system for using remote sensing techniques to detect and evaluate air pollution effects on forest stands. In: Sixth International Symposium on Remote Sensing of Environment, Vol. II, Proceedings, pp. 1169 - 1173.

Winterberger, K.C. 1984. LANDSAT data and aerial photographs used in a multiphase sample of vegetation and related resources in interior Alaska. In: Inventorying Forest and other Vegetation of the High Latitude High Altitude Regions. Proceedings of an International Symposium. Society of American Foresters, Regional Technical Conference, 23-26, July, Fairbanks, Alaska, pp. 157 - 163.

Yandle, D.O.; White, F.M. 1977. An application of two stage forest sampling. In: Southern Journal of Applied Forestry, 1(3):27 - 32.